

```

131 AAAAGGAGAGTATGATGATTAACATTTCCGCTGTCGCCCTTATTCCCTCTTTTTCGGCA 190
191 TTTTGGCTTCTCTTTTGTCTACCCAGAAAGCGTGTCAAGTAAAGATGCTGAAGAT 250
191 TTTTGGCTTCTCTTTTGTCTACCCAGAAAGCGTGTCAAGTAAAGATGCTGAAGAT 250
251 CAGTTGGGTCACGAGTGGGTATCATCGAAGTGGATCTCAACAGCGGTGAAGATCTTGG 310
251 CAGTTGGGTCACGAGTGGGTATCATCGAAGTGGATCTCAACAGCGGTGAAGATCTTGG 310
311 AGTTTTCGCCCGCAAGACGTTCTCCAATGATGAGCACATTTTAAAGTTCCTGATGTCGC 370
311 AGTTTTCGCCCGCAAGACGTTTTCATGATGAGCACATTTTAAAGTTCCTGATGTCGC 370
371 GCGGTATATPCCGTTGTAGCGCGGCAAGAGCAACTCGGTGCGCGCATACACTATTCT 430
371 GCGGTATATPCCGTTGTAGCGCGGCAAGAGCAACTCGGTGCGCGCATACACTATTCT 430
431 CAGAAATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCAATCTTACGATGCGATGACA 490
431 CAGAAATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCAATCTTACGATGCGATGACA 490
491 GTAAGAGAAAT 500
491 GTAAGAGAAAT 500

```

```

RESULT 15
LOCUS CD281010 582 bp mRNA linear EST 23-MAY-2003
DEFINITION G44224.52 NCI_CGAP_Zemb3 Danio rerio cDNA clone IMAGE:6320701 5',
mRNA sequence.
ACCESSION CD281010
VERSION CD281010.1 GI:31058786
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
Authors: Arundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
Expressed sequence tags from NCI_CGAP_Zemb3, a Danio rerio
embryonic library
Unpublished (2003)
Contact: Chen F.
Exelixis, Inc.
170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA Sequencing by: Exelixis, Inc. Clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 13759 row: H column: 14
High quality sequence stop: 582.
Location/Qualifiers
1..582
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6320701"
/tissue_type="embryo"
/lab_host="DH10B (Tl-resistant)"
/clone_lib="NCI_CGAP_Zemb3"
/note="Vector: PCMV-SPOrt6.1; Site1: EcoRV; Site 2: NotI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 Kb. Constructed by J. Wang (Research Genetics,
Invitrogen Corp) from tissue donated by L. Zon (Harvard
University). Note: this is a NCI_CGAP Library."

```

FEATURES

source

```

Query Match 97.0%; Score 485.2; DB 14; Length 582;
Best Local Similarity 99.4%; Pred. No. 1.9e-117;
Matches 487; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 11 TCAGTGGCACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATTTCTAAATA 70
Db 8 TTAGTGGCACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATTTCTAAATA 67
Qy 71 CATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATATGGA 130
Db 68 CATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATATGGA 127
Qy 131 AAAAGGAGAGTATGATGATTAACATTTCCGCTGTCGCCCTTATTCCCTCTTTTTCGGCA 190
Db 128 AAAAGGAGAGTATGATGATTAACATTTCCGCTGTCGCCCTTATTCCCTCTTTTTCGGCA 187
Qy 191 TTTTGGCTTCTCTTTTGTCTACCCAGAAAGCGTGTCAAGTAAAGATGCTGAAGAT 250
Db 188 TTTTGGCTTCTCTTTTGTCTACCCAGAAAGCGTGTCAAGTAAAGATGCTGAAGAT 247
Qy 251 CAGTTGGGTCACGAGTGGGTATCATCGAAGTGGATCTCAACAGCGGTGAAGATCTTGG 310
Db 248 CAGTTGGGTCACGAGTGGGTATCATCGAAGTGGATCTCAACAGCGGTGAAGATCTTGG 307
Qy 311 AGTTTTCGCCCGCAAGACGTTCTCCAATGATGAGCACATTTTAAAGTTCCTGATGTCGC 370
Db 308 AGTTTTCGCCCGCAAGACGTTTTCATGATGAGCACATTTTAAAGTTCCTGATGTCGC 367
Qy 371 GCGGTATATPCCGTTGTAGCGCGGCAAGAGCAACTCGGTGCGCGCATACACTATTCT 430
Db 368 GCGGTATATPCCGTTGTAGCGCGGCAAGAGCAACTCGGTGCGCGCATACACTATTCT 427
Qy 431 CAGAAATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCAATCTTACGATGCGATGACA 490
Db 428 CAGAAATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCAATCTTACGATGCGATGACA 487
Qy 491 GTAAGAGAAAT 500
Db 488 GTAAGAGAAAT 497

```

Search completed: April 22, 2004, 12:56:08
Job time : 3057.17 secs

```

ACCESSION      CD280954
VERSION        CD280954.1
KEYWORDS       EST.
SOURCE         Danio rerio (zebrafish)
ORGANISM       Danio rerio

REFERENCE
AUTHORS        Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
               Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
TITLE          Expressed sequence tags from NCI_CGAP_Zemb3, a Danio rerio
               embryonic library
JOURNAL        Unpublished (2003)
COMMENT        Contact: Chen F.
               Exelixis, Inc.
               170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
               Tel: 650 837 7000
               Fax: 650 837 8300
               Email: fchen@exelixis.com
               DNA Sequencing by: Exelixis, Inc. Clone distribution information
               can be found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: 13759 row: E column: 24
               High quality sequence stop: 572.
               Location/Qualifiers
                 1..572
                   /organism="Danio rerio"
                   /mol_type="mRNA"
                   /db_xref="taxon:7955"
                   /clone="IMAGE:6320639"
                   /tissue_type="embryo"
                   /lab_host="DH10B (TI-resistant)"
                   /clone_lib="NCI CGAP Zemb3"
                   /note="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI;
                   Cloned unidirectionally. Primer: Oligo dT. Average insert
                   size 2.1 kb. Constructed by J. Wang (Research Genetics,
                   Invitrogen Corp) from tissue donated by L. Zon (Harvard
                   University). Note: this is a NCI_CGAP Library."

FEATURES
source
1..572
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:538359"
/db_xref="taxon:112509"
/clone="HP10C15"
/tissue_type="epidermis"
/dev_stage="seedlings grown in greenhouse for 6 days"
/lab_host="XL10-Gold"
/clone_lib="Hp"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
cDNA); Site 2: XhoI (3'-end of cDNA). Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable. Average insert size is
850 bp"

ORIGIN
Query Match      97.0%; Score 485.2; DB 14; Length 572;
Best Local Similarity 99.4%; Pred. No. 1.9e-117;
Matches 487; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 TCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAAATA 70
DB 11 TCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAAATA 70
QY 71 CATTCAAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATTGA 130
DB 71 CATTCAAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATTGA 130
QY 131 AAAGAGAGATGTAGTATTCAACATTTCCGCTGCGCCCTATTCCCTTTTGGGCA 190
DB 131 AAAGAGAGATGTAGTATTCAACATTTCCGCTGCGCCCTATTCCCTTTTGGGCA 190
QY 166 AAAGAGAGATGTAGTATTCAACATTTCCGCTGCGCCCTATTCCCTTTTGGGCA 225
DB 166 AAAGAGAGATGTAGTATTCAACATTTCCGCTGCGCCCTATTCCCTTTTGGGCA 225
QY 191 TTTTCCTCTTCCTGTTTCTCACCAGAAAGCTGTGGAAGTAAAGATGCTGAAGAT 250
DB 191 TTTTCCTCTTCCTGTTTCTCACCAGAAAGCTGTGGAAGTAAAGATGCTGAAGAT 250
QY 226 TTTTCCTCTTCCTGTTTCTCACCAGAAAGCTGTGGAAGTAAAGATGCTGAAGAT 285
DB 226 TTTTCCTCTTCCTGTTTCTCACCAGAAAGCTGTGGAAGTAAAGATGCTGAAGAT 285
QY 251 CAGTTGGGTGCAGAGTGGTGTACATCGAATGATCTCAACAGCGGTAAAGATCCTTGA 310
DB 251 CAGTTGGGTGCAGAGTGGTGTACATCGAATGATCTCAACAGCGGTAAAGATCCTTGA 310
QY 286 CAGTTGGGTGCAGAGTGGTGTACATCGAATGATCTCAACAGCGGTAAAGATCCTTGA 345
DB 286 CAGTTGGGTGCAGAGTGGTGTACATCGAATGATCTCAACAGCGGTAAAGATCCTTGA 345
QY 311 AGTTTTCGCCCGGAGAGAGCTTCTCCATGATGAGACATTTTAAAGTCTGCTATGGC 370
DB 311 AGTTTTCGCCCGGAGAGAGCTTCTCCATGATGAGACATTTTAAAGTCTGCTATGGC 370
QY 346 AGTTTTCGCCCGGAGAGAGCTTCTCCATGATGAGACATTTTAAAGTCTGCTATGGC 405
DB 346 AGTTTTCGCCCGGAGAGAGCTTCTCCATGATGAGACATTTTAAAGTCTGCTATGGC 405
QY 371 GCGGTATTATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGGATACACTATCT 430
DB 371 GCGGTATTATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGGATACACTATCT 430
QY 406 GCGGTATTATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGGATACACTATCT 465
DB 406 GCGGTATTATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGGATACACTATCT 465

QY 431 CAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGCCATGACA 490
DB 431 CAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGCCATGACA 490
QY 466 CAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGCCATGACA 525
DB 466 CAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGCCATGACA 525

QY 491 GTAAGAGAAAT 500
DB 491 GTAAGAGAAAT 535

RESULT 14
CB878888
LOCUS      HP10C15 HP Hordeum vulgare subsp. vulgare cDNA clone HP10C15
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION  CB878888
VERSION    CB878888.1
KEYWORDS   Hordeum vulgare subsp. vulgare
ORGANISM   Hordeum vulgare subsp. vulgare
REFERENCE  Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
           EST sequencing and analysis in barley (2002)
           Unpublished (2002)
           Contact: Stein Nils
           Molecular Markers Group, Department Genbank
           Institute of Plant Genetics and Crop Plant Research (IPK)
           Corrensstr. 3 06466, Gatersleben, Germany
           Tel: 039482-5522
           Fax: 039482-5595
           Email: stein@ipk-gatersleben.de
           Insert Length: 578 Std Error: 0.00
           Plate: 10 row: C column: 15
           Seq primer: T3
           Location/Qualifiers
             1..578
               /organism="Hordeum vulgare subsp. vulgare"
               /mol_type="mRNA"
               /cultivar="barke"
               /sub_species="vulgare"
               /db_xref="GABI:538359"
               /db_xref="taxon:112509"
               /clone="HP10C15"
               /tissue_type="epidermis"
               /dev_stage="seedlings grown in greenhouse for 6 days"
               /lab_host="XL10-Gold"
               /clone_lib="Hp"
               /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
               cDNA); Site 2: XhoI (3'-end of cDNA). Due to a cloning
               artefact caused by the kit, in most cases the EcoRI site
               is NOT present, as well as the EcoRI adapter used for
               cloning. To excise the insert, restriction sites upstream
               EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
               due to the cloning system used Blue/white selection for
               recombinants is not 100% reliable. Average insert size is
               850 bp"

```



```

Db      372  GGGTATTATCCGGTATTGACCGCGGCAAGAGCACTCGTTCGGCGCATACACTATTCT 431
Qy      431  CAGATGACTTGGTGGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACA 490
Db      432  CAGATGACTTGGTGGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACA 491
Qy      491  GTAAGAGAAT 500
Db      492  GTAAGAGAAT 501

RESULT 10
CB878190
LOCUS
DEFINITION
  HP07K12 HP Hordeum vulgare subsp. vulgare cDNA clone HP07K12
  5-PRIME, mRNA sequence.
ACCESSION
  CB878190
VERSION
  CB878190.1 GI:30080176
SOURCE
  EST.
  Hordeum vulgare subsp. vulgare
  Hordeum vulgare subsp. vulgare
  Hordeum vulgare subsp. vulgare
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Poideae; Triticeae; Hordeum.
  1 (bases 1 to 540)
  Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
  EST sequencing and analysis in barley (2002)
  Unpublished (2002)
  Contact: Stein Nils
  Molecular Markers Group, Department Genbank
  Institute of Plant Genetics and Crop Plant Research (IPK)
  Corrensstr. 3 06466, Gatersleben, Germany
  Tel: 039482-5522
  Fax: 039482-5595
  Email: stein@ipk-gatersleben.de
  Insert Length: 540 Std Error: 0.00
  Plate: 7 row: K column: 12
  Seq primer: T3.
  Location/Qualifiers
    1..540
      /organism="Hordeum vulgare subsp. vulgare"
      /mol_type="mRNA"
      /cultivar="barke"
      /sub_species="vulgare"
      /db_xref="GABI:537645"
      /db_xref="taxon:112509"
      /clone="HP07K12"
      /tissue_type="epidermis"
      /dev_stage="seedlings grown in greenhouse for 6 days"
      /lab_host="Xl10-Gold"
      /clone_lib="Hp"
      /note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
      cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning
      artefact caused by the kit, in most cases the EcoRI site
      is NOT present, as well as the EcoRI adapter used for
      cloning. To excise the insert, restriction sites upstream
      EcoRI should be used (e.g. BamHI, Sali, PstI). NOTE: Also
      due to the cloning system used Blue/white selection for
      recombinants is not 100% reliable. Average insert size is
      850 bp"

ORIGIN
Query Match      97.0%; Score 485.2; DB 14; Length 540;
Best Local Similarity 99.4%; Pred. No. 1.9e-117;
Matches 487; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      11  TCAGGTGGCACTTTTCGGGGAATGCGCGGAACCCCTATTGTTATTTTCTAAATA 70
Db      13  TTAGTGGCACTTTTCGGGGAATGCGCGGAACCCCTATTGTTATTTTCTAAATA 72
Qy      71  CATTCAAAATGATGATCCGCTCATGACAAATACCCCTGATAAATGCTTCAATAATTGA 130

```

```

Db      73  CATTCAAAATGATGATCCGCTCATGACAAATACCCCTGATAAATGCTTCAATAATTGA 132
Qy      131  AAAGGAAGATGATGATTAATCAACATTTCCGTGTCGCCCTTATTCCTTTTGGCGCA 190
Db      133  AAAGGAAGATGATGATTAATCAACATTTCCGTGTCGCCCTTATTCCTTTTGGCGCA 192
Qy      191  TTTTGCCTTCCTGTTTGTCTACCCAGAAACGCTGGTGAAGTAAAGTAAAGATGCTGAAGAT 250
Db      193  TTTTGCCTTCCTGTTTGTCTACCCAGAAACGCTGGTGAAGTAAAGTAAAGATGCTGAAGAT 252
Qy      251  CAGTGGGTGCACAGTGGGTATACATCGAATCGAATCTCAACAGCGGTAAGATCCTTGAG 310
Db      253  CAGTGGGTGCACAGTGGGTATACATCGAATCGAATCTCAACAGCGGTAAGATCCTTGAG 312
Qy      311  AGTTTTCGCCGCCCGGAGAACGTTCTCCAATGATGAGCACATTTTAAAGTCTTGCTATGGC 370
Db      313  AGTTTTCGCCGCCCGGAGAACGTTTTCOAATGATGAGCACATTTTAAAGTCTTGCTATGGC 372
Qy      371  GCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGTCCGCGCATACACTATTCT 430
Db      373  GCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGTCCGCGCATACACTATTCT 432
Qy      431  CAGAATGACTTGGTGTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACA 490
Db      433  CAGAATGACTTGGTGTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACA 492
Qy      491  GTAAGAGAAT 500
Db      493  GTAAGAGAAT 502

RESULT 11
BU998674
LOCUS
DEFINITION
  BU998674 HI Hordeum vulgare subsp. vulgare cDNA clone HillN02
  5-PRIME, mRNA sequence.
ACCESSION
  BU998674
VERSION
  BU998674.1 GI:24275657
KEYWORDS
  EST.
SOURCE
  Hordeum vulgare subsp. vulgare
  Hordeum vulgare subsp. vulgare
  Hordeum vulgare subsp. vulgare
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Poideae; Triticeae; Hordeum.
  1 (bases 1 to 567)
  Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
  EST sequencing and analysis in barley (2002)
  Unpublished (2002)
  Contact: Stein Nils
  Molecular Markers Group, Department Genbank
  Institute of Plant Genetics and Crop Plant Research (IPK)
  Corrensstr. 3 06466, Gatersleben, Germany
  Tel: 039482-5522
  Fax: 039482-5595
  Email: stein@ipk-gatersleben.de
  Insert Length: 567 Std Error: 0.00
  Plate: 11 row: N column: 2
  Seq primer: M13rev.
  Location/Qualifiers
    1..567
      /organism="Hordeum vulgare subsp. vulgare"
      /mol_type="mRNA"
      /cultivar="barke"
      /sub_species="vulgare"
      /db_xref="GABI:252538"
      /db_xref="taxon:112509"
      /clone="HillN02"
      /tissue_type="female inflorescences"
      /dev_stage="female inflorescences (approx. 3 mm in size)"
      /lab_host="Xl10-Gold"
      /clone_lib="Hi"
      /notes="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of
      cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning

```


Exelixis, Inc.
170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA sequencing by: Exelixis, Inc. Clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 13759 row: H column: 2
High quality sequence stop: 518.

FEATURES

Location/Qualifiers
1. 518
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6320689"
/tissue_type="embryo"
/lab_host="DH10B (T1-resistant)"
/clone_lib="NCI_CGAP_ZEM3"
/notes="vector: PCMV-Spork1.1; Site_1: EcoRV; Site_2: NotI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by J. Wang (Research Genetics,
Invitrogen Corp) from tissue donated by L. Zon (Harvard
University). Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 97.0%; Score 485.2; DB 14; Length 518;
Best Local Similarity 99.4%; Pred. No. 1.9e-117;
Matches 487; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 TCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTATTTTCTAAATA 70
DB 21 TTAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTATTTTCTAAATA 80

QY 71 CATTCAAATATGATCGGCTCATGAGACATAACCTGTAAATGCTTCAATAATTGA 130
DB 81 CATTCAAATATGATCGGCTCATGAGACATAACCTGTAAATGCTTCAATAATTGA 140

QY 131 AAAAGGAAGATGATGAGTATTCACATTTCCGTGCGCCCTATTCCCTTTTGGCGCA 190
DB 141 AAAAGGAAGATGAGTATTCACATTTCCGTGCGCCCTATTCCCTTTTGGCGCA 200

QY 191 TTTTGCCTTCCTGTTTGTCTACCCAGAAACCGCTGGTGAAGTAAAGATGCTGAAGAT 250
DB 201 TTTTGCCTTCCTGTTTGTCTACCCAGAAACCGCTGGTGAAGTAAAGATGCTGAAGAT 260

QY 251 CAGTTGGTGACGAGTGGGTATCATCGAATCGATCTCAACAGCGGTAAAGTCCCTGAG 310
DB 261 CAGTTGGTGACGAGTGGGTATCATCGAATCGATCTCAACAGCGGTAAAGTCCCTGAG 320

QY 311 AGTTTCGCCCCGGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCGTATGTCG 370
DB 321 AGTTTCGCCCCGGAAGAACGTTTCCAATGATGAGCACTTTTAAAGTTCGTATGTCG 380

QY 371 GCGGTATATCCCGTGTGAGCGGGGCAAGACCACTGGTGGCGGCATACACTATTCT 430
DB 381 GCGGTATATCCCGTGTGAGCGGGGCAAGACCACTGGTGGCGGCATACACTATTCT 440

QY 431 CAGAAATGACTTGGTGTAGTACTACCCAGTCAAGAAAGCATCTTACCGATGCGCATGACA 490
DB 441 CAGAAATGACTTGGTGTAGTACTACCCAGTCAAGAAAGCATCTTACCGATGCGCATGACA 500

QY 491 GTAAGAGAAAT 500
DB 501 GTAAGAGAAAT 510

RESULT 9
BU993519 535 bp mRNA linear EST 22-OCT-2002
LOCUS HD13M20r HD Hordeum vulgare cDNA clone HD13M20 5-PRIME, mRNA
DEFINITION
sequence.
ACCESSION BU993519

VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL COMMENT

BU993519.1 GI:24244465
EST.
Hordeum vulgare
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 535)
Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
EST sequencing and analysis in barley (2002)
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 535 Std Error: 0.00
Plate: 13 row: M column: 20
Seq primer: M13rev.
Location/Qualifiers
1. 535
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Golden Promise"
/db_xref="GABI:243631"
/db_xref="taxon:4513"
/clone="HD13M20"
/tissue_type="callus"
/dev_stage="callus (5-10 mm in diameter)"
/lab_host="XL10-Gold"
/clone_lib="HD"
/note="vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
artifact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable. Average insert size is 1
kb"

FEATURES

Location/Qualifiers
1. 535
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Golden Promise"
/db_xref="GABI:243631"
/db_xref="taxon:4513"
/clone="HD13M20"
/tissue_type="callus"
/dev_stage="callus (5-10 mm in diameter)"
/lab_host="XL10-Gold"
/clone_lib="HD"
/note="vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
artifact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable. Average insert size is 1
kb"

ORIGIN

Query Match 97.0%; Score 485.2; DB 13; Length 535;
Best Local Similarity 99.4%; Pred. No. 1.9e-117;
Matches 487; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 TCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTATTTTCTAAATA 70
DB 12 TTAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTATTTTCTAAATA 71

QY 71 CATTCAAATATGATCGGCTCATGAGACATAACCCCTGATAAATGCTTCAATAATTGA 130
DB 72 CATTCAAATATGATCGGCTCATGAGACATAACCCCTGATAAATGCTTCAATAATTGA 131

QY 131 AAAAGGAAGATGAGTATTCACATTTCCGTGCGCCCTATTCCCTTTTGGCGCA 190
DB 132 AAAAGGAAGATGAGTATTCACATTTCCGTGCGCCCTATTCCCTTTTGGCGCA 191

QY 191 TTTTGCCTTCCTGTTTGTCTACCCAGAAACCGCTGGTGAAGTAAAGATGCTGAAGAT 250
DB 192 TTTTGCCTTCCTGTTTGTCTACCCAGAAACCGCTGGTGAAGTAAAGATGCTGAAGAT 251

QY 251 CAGTTGGTGACGAGTGGGTATCATCGAATCGATCTCAACAGCGGTAAAGTCCCTGAG 310
DB 252 CAGTTGGTGACGAGTGGGTATCATCGAATCGATCTCAACAGCGGTAAAGTCCCTGAG 311

QY 311 AGTTTCGCCCCGGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCGTATGTCG 370
DB 312 AGTTTCGCCCCGGAAGAACGTTTCCAATGATGAGCACTTTTAAAGTTCGTATGTCG 371

QY 371 GCGGTATATCCCGTGTGAGCGGGGCAAGACCACTGGTGGCGGCATACACTATTCT 430

```

/clone="msh2_3073"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."
ORIGIN
Query Match          97.1%; Score 485.6; DB 28; Length 773;
Best Local Similarity 99.2%; Pred. No. 1.5e-117;
Matches 488; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 9 CGTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTTCTAAA 68
Db 81 CTTTAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTTCTAAA 140
QY 69 TACATTCAATATGATTCGGCTCATGAGACAATAACCCCTGATAATGCTTCAATAATTT 128
Db 141 TACATTCAATATGATTCGGCTCATGAGACAATAACCCCTGATAATGCTTCAATAATTT 200
QY 129 GAAAAGGAAGAGTATGAGTATTCACATTTCCGCTGCGCCCTATTCCCTTTTGGCG 188
Db 201 GAAAAGGAAGAGTATGAGTATTCACATTTCCGCTGCGCCCTATTCCCTTTTGGCG 260
QY 189 CATTTTGCCTTCCTGTTTTCCTACCCGAAACCGTGTGGAAGTAAAGATGCTGAAG 248
Db 261 CATTTTGCCTTCCTGTTTTCCTACCCGAAACCGTGTGGAAGTAAAGATGCTGAAG 320
QY 249 ATCAGTTGGGTGCACGAGTGGTTCATCGAACTGATCTCAACAGCGGTAAAGTCCCTTG 308
Db 321 ATCAGTTGGGTGCACGAGTGGTTCATCGAACTGATCTCAACAGCGGTAAAGTCCCTTG 380
QY 309 AGATTTTTCGCCCGAAGACGTTTCCAAATGATGAGCACTTTTAAAGTCTGCTATGTG 368
Db 381 AGATTTTTCGCCCGAAGACGTTTCCAAATGATGAGCACTTTTAAAGTCTGCTATGTG 440
QY 369 GCGCGGTATATCCGCTGTTGACGCGGCAAGCACTCGTCCGCGCATACACTATT 428
Db 441 GCGCGGTATATCCGCTGTTGACGCGGCAAGCACTCGTCCGCGCATACACTATT 500
QY 429 CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA 488
Db 501 CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA 560
QY 489 CAGTAAGAGAAT 500
Db 561 CAGTAAGAGAAT 572

RESULT 7
BZ573360
LOCUS
DEFINITION
msh2_3069.y2 msh Pseudomonas aeruginosa genomic clone msh2_3069,
genomic survey sequence.
ACCESSION
BZ573360
VERSION
BZ573360.1 GI:27208421
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 781)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

```

FEATURES

source

Location/Qualifiers

```

1..781
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone_lib="msh2_3069"
/note="Environmental isolate. Whole genomic shotgun
library."
ORIGIN
Query Match          97.1%; Score 485.6; DB 28; Length 781;
Best Local Similarity 99.2%; Pred. No. 1.5e-117;
Matches 488; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 9 CGTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTTCTAAA 68
Db 82 CTTTAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTTCTAAA 141
QY 69 TACATTCAATATGATTCGGCTCATGAGACAATAACCCCTGATAATGCTTCAATAATTT 128
Db 142 TACATTCAATATGATTCGGCTCATGAGACAATAACCCCTGATAATGCTTCAATAATTT 201
QY 129 GAAAAGGAAGAGTATGAGTATTCACATTTCCGCTGCGCCCTATTCCCTTTTGGCG 188
Db 202 GAAAAGGAAGAGTATGAGTATTCACATTTCCGCTGCGCCCTATTCCCTTTTGGCG 261
QY 189 CATTTTGCCTTCCTGTTTTCCTACCCGAAACCGTGTGGAAGTAAAGATGCTGAAG 248
Db 262 CATTTTGCCTTCCTGTTTTCCTACCCGAAACCGTGTGGAAGTAAAGATGCTGAAG 321
QY 249 ATCAGTTGGGTGCACGAGTGGTTCATCGAACTGATCTCAACAGCGGTAAAGATGCTTG 308
Db 322 ATCAGTTGGGTGCACGAGTGGTTCATCGAACTGATCTCAACAGCGGTAAAGATGCTTG 381
QY 309 AGATTTTTCGCCCGAAGACGTTTCCAAATGATGAGCACTTTTAAAGTCTGCTATGTG 368
Db 382 AGATTTTTCGCCCGAAGACGTTTCCAAATGATGAGCACTTTTAAAGTCTGCTATGTG 441
QY 369 GCGCGGTATATCCGCTGTTGACGCGGCAAGCACTCGTCCGCGCATACACTATT 428
Db 442 GCGCGGTATATCCGCTGTTGACGCGGCAAGCACTCGTCCGCGCATACACTATT 501
QY 429 CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA 488
Db 502 CTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA 561
QY 489 CAGTAAGAGAAT 500
Db 562 CAGTAAGAGAAT 573

RESULT 8
CD280998
LOCUS
DEFINITION
G42424.4 NCI_CGAP_ZEmb3 Danio rerio cDNA clone IMAGE:6320689 5',
mRNA sequence.
ACCESSION
CD280998
VERSION
CD280998.1 GI:31058774
KEYWORDS
EST.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
REFERENCE
1 (bases 1 to 518)
Murray,L., Oliva,J., Park,C., Reyes,J., Yungen,J. and Swimmer,C.
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
Amundsen,C., Cachuela,N., Chen,F., Cheung,L.M., Chong,A.,
Murray,L., Oliva,J., Park,C., Reyes,J., Yungen,J. and Swimmer,C.
Expressed sequence tags from NCI_CGAP_ZEmb3, a Danio rerio
embryonic library
Unpublished (2003)
Contact: Chen F.

```

Db 3630 AAAGGAAGATGATGATTAACACATTCCTGTCGCCCTTATCCCTTTTGGCGAT 3689
 Qy 192 TTTCCTTCCTGTTTGTCTACCCAGAAACCTGCTGTAAGATGCTGAAGATC 251
 Db 3690 TTTCCTTCCTGTTTGTCTACCCAGAAACCTGCTGTAAGATGCTGAAGATC 3749
 Qy 252 AGTTGGGTCAGAGTGGGTTTACATCGAATCTGATCTCAACAGCGGTGAAGATCCTTGAGA 311
 Db 3750 AGTTGGGTCAGAGTGGGTTTACATCGAATCTGATCTCAACAGCGGTGAAGATCCTTGAGA 3809
 Qy 312 GTTTTCGCCGCCGGAAGAACTTCTCAATGATGAGACATTTTAAAGTTCCTGCTATGTCGCG 371
 Db 3810 GTTTTCGCCGCCGGAAGAACTTCTCAATGATGAGACATTTTAAAGTTCCTGCTATGTCGCG 3869
 Qy 372 CGGTATATTCCTGTTTGAAGCGGCGCAAGAGCACTCGTTCGCGCGCATACATATTC 431
 Db 3870 CGGTATATTCCTGTTTGAAGCGGCGCAAGAGCACTCGTTCGCGCGCATACATATTC 3929
 Qy 432 AGAATGACTTGGTTGAGTACTCACCAGTCCAGAAAGCATCTTACGGATGCGATGACAG 491
 Db 3930 AGAATGACTTGGTTGAGTACTCACCAGTCCAGAAAGCATCTTACGGATGCGATGACAG 3989
 Qy 492 TAAGAGAAT 500
 Db 3990 TAAGAGAAT 3998

RESULT 5
 CD280127
 LOCUS
 DEFINITION
 CD280127.1 GI:31057903
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Danio rerio (zebrafish)
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
 Murray, B., Oliveira, J., Park, C., Reyes, J., Fungen, J. and Swimmer, C.
 Expressed sequence tags from NCI_CGAP_Zemb3, a Danio rerio
 embryonic library
 Unpublished (2003)
 Contact: Chen F.
 Exelixis, Inc.
 170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
 Tel: 650 837 7000
 Fax: 650 837 8300
 Email: fchen@exelixis.com
 DNA Sequencing by: Exelixis, Inc. Clone distribution information
 can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: 13762 row: F column: 22
 High quality sequence stop: 663.
 Location/Qualifiers
 1. .663
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:6321813"
 /tissue_type="embryo"
 /lab_host="DH10B (Tl-resistant)"
 /clone_lib="NCI_CGAP_Zemb3"
 /notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 kb. Constructed by J. Wang (Research Genetics,
 Invitrogen Corp) from tissue donated by L. Zon (Harvard
 University). Note: this is a NCI_CGAP Library."

FEATURES
 source
 1. .663
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:6321813"
 /tissue_type="embryo"
 /lab_host="DH10B (Tl-resistant)"
 /clone_lib="NCI_CGAP_Zemb3"
 /notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 kb. Constructed by J. Wang (Research Genetics,
 Invitrogen Corp) from tissue donated by L. Zon (Harvard
 University). Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 97.1%; Score 485.6; DB 14; Length 663;
 Best Local Similarity 98.2%; Pred. No. 1.5e-117;
 Matches 491; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 GTTAACTACGTCAGGTGGACATTTTCGGGGAATGTCGGGAACCCCTATTGTTTATT 60
 Db 87 GCTTAACTATAGGTGGACATTTTCGGGGAATGTCGGGAACCCCTATTGTTTATT 146
 Qy 61 TTTCTAAATACATTTCAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
 Db 147 TTTCTAAATACATTTCAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 206
 Qy 121 ATATATTTGAAAAGGAAGATGATGATTTCAACATTTCCGTCGCGCTTATTCCTT 180
 Db 207 ATATATTTGAAAAGGAAGATGATGATTTCAACATTTCCGTCGCGCTTATTCCTT 266
 Qy 181 TTTTGGCGGCAATTTTGCCTTCTCTCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTT 240
 Db 267 TTTTGGCGGCAATTTTGCCTTCTCTCTTTCCTTTCCTTTCCTTTCCTTTCCTT 326
 Qy 241 TGCTGAAGATCAGTTGGTGCACAGAGTGGTGTACATCGAACTGGATCTCAACAGCGGTAA 300
 Db 327 TGCTGAAGATCAGTTGGTGCACAGAGTGGTGTACATCGAACTGGATCTCAACAGCGGTAA 386
 Qy 301 GATCCTTTGAGAGTTTTCGCCCGGAAAGCGTTCTCCATGATGAGCACTTTTAAAGTTCT 360
 Db 387 GATCCTTTGAGAGTTTTCGCCCGGAAAGCGTTTTCATGATGAGCACTTTTAAAGTTCT 446
 Qy 361 GCTATGTGGCGCGGTATATTCCTGTTTTCGCGCGGGAAGAGCAACTCGTTCGCGCAT 420
 Db 447 GCTATGTGGCGCGGTATATTCCTGTTTTCGCGCGGGAAGAGCAACTCGTTCGCGCAT 506
 Qy 421 ACATATTTCTCAGATGACTTGGTGTGAGTACTCACCAGTCCAGAAAGCATCTTACCGGA 480
 Db 507 ACATATTTCTCAGATGACTTGGTGTGAGTACTCACCAGTCCAGAAAGCATCTTACCGGA 566
 Qy 481 TGGCATGACAGTAAGAGAAT 500
 Db 567 TGGCATGACAGTAAGAGAAT 586

RESULT 6
 B2573370
 LOCUS
 DEFINITION
 msh2_3073.y2 msh Pseudomonas aeruginosa genomic clone msh2_3073,
 genomic survey sequence.
 ACCESSION
 B2573370
 VERSION
 B2573370.1 GI:27208431
 KEYWORDS
 GSS.
 SOURCE
 Pseudomonas aeruginosa
 Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE
 1 (bases 1 to 773)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) In press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: Shotgun.
 Location/Qualifiers
 1. .773
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="MSH"
 /db_xref="taxon:287"

```

Db      477 ATTCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAGCATCTTACGGATGCA 536
QY      486 TGACAGTAAGAGAT 500
Db      537 TGACAGTAAGAGAT 551

RESULT 3
BQ817108
LOCUS      660 bp mRNA linear EST 01-AUG-2002
DEFINITION Lambda Zap II Chlamydomonas reinhardtii CC-1690, Deflagellation (normalized),
ACCESSION BQ817108
VERSION    BQ817108.1 GI:22066733
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii
ORGANISM   Chlamydomonas reinhardtii
REFERENCE  1 (bases 1 to 660)
AUTHORS    Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,
           Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.
TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model
           Unicellular System for Analyzing Gene Function and Regulation in
           Vascular Plants. Project: 1030
JOURNAL    Unpublished (2002)
COMMENT    Contact: Charles Hauser
           DCMB Box 91000
           Duke University
           Durham, NC 27708-1000
           Tel: 919 613 8159
           Fax: 919 613 8177
           Email: chauser@duke.edu.

FEATURES   source
            1..660
               /organism="Chlamydomonas reinhardtii"
               /mol_type="mRNA"
               /strain="CC-1690 wild type mt+ 21gr"
               /db_xref="taxon:3055"
               /clone_lib="C. reinhardtii CC-1690, Deflagellation
               (normalized), Lambda Zap II"
               /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
               XhoI; Deflagellation library, constructed by John Davies
               and Jeffrey McPherson, combines cDNAs from CC-1690 cells
               which had been re-synthesizing flagella for 15, 30 and 60
               min after being deflagellated by pH shock. PolyA mRNA was
               purified from each sample, pooled and cDNA synthesized.
               The cDNA was directionally cloned into lambda Zap II
               (Stratagene) in the EcoRI (5') and XhoI (3') sites.
               pBluescript II SK- plasmids were excised from the lambda
               Zap clones by superinfection with ExAssist (Stratagene)
               phage. The library was normalized using method 4 described
               in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN
Query Match      97.2%; Score 485.8; DB 13; Length 660;
Best Local Similarity 99.6%; Pred. No. 1.3e-117;
Matches 487; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      9 CGTCAGTGGCACTTTTCGGGAAATGCGCGGAACCCCTATTGTTTATTCTTCTAAA 68
Db      172 CGTCAGTGGCACTTTTCGGGAAATGCGCGGAACCCCTATTGTTTATTCTTCTAAA 231

QY      69 TACATTCAATATGATCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATAATATT 128
Db      232 TACATTCAATATGATCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATAATATT 291

QY      129 GAAAAAGGAAGATATGAGTATCAACATTCGGTGTGCGCCCTATTTCCTTTTGGGG 188
Db      292 GAAAAAGGAAGATATGAGTATCAACATTCGGTGTGCGCCCTATTTCCTTTTGGGG 351

QY      189 CATTTTGCCTTCTCTGTTTGTCTACCCAGAAACGCTGGTGAAGATAAAGATGCTGAAG 248

```

```

Db      352 CATTTTGCCTTCTCTGTTTGTCTACCCAGAAACGCTGGTGAAGATAAAGATGCTGAAG 411
QY      249 ATCAGTTGGGTGCACGAGTGGGTTCATCGAACTGGATCTCAACAGCGGTAAAGATCCTTG 308
Db      412 ATCAGTTGGGTGCACGAGTGGGTTCATCGAACTGGATCTCAACAGCGGTAAAGATCCTTG 471
QY      309 AGAGTTTTCGCCGCCGGAAGAACGTTTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGTG 368
Db      472 AGAGTTTTCGCCGCCGGAAGAACGTTTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGTG 531
QY      369 GCGCGGTATTATCCCGTGTGACGCGGGCAAGAGCAATCGGTGCGCCGATACATATT 428
Db      532 GCGCGGTATTATCCCGTGTGACGCGGGCAAGAGCAATCGGTGCGCCGATACATATT 591
QY      429 CTCAGATGACTGTGTTGAGTACTACCAAGTACACAGAAAGCATCTTACGGATGGCATGA 488
Db      592 CTCAGATGACTGTGTTGAGTACTACCAAGTACACAGAAAGCATCTTACGGATGGCATGA 651
QY      489 CAGTAAGAG 497
Db      652 CAGTAAGAG 660

RESULT 4
CF753623
LOCUS      4096 bp mRNA linear EST 16-OCT-2003
DEFINITION EST-Contig462 Preamplified custom cDNA library in pMAB58 (ResGen,
ACCESSION CF753623
VERSION    CF753623.1 GI:37697416
KEYWORDS   EST.
SOURCE     Emiliana huxleyi
ORGANISM   Emiliana huxleyi
REFERENCE  1 (bases 1 to 4096)
AUTHORS    Wahlund, T.M., Hadaegh, A.R., Clark, R., Nguyen, B., Fanelli, M. and
           Read, B.
TITLE      Analysis of Expressed Sequence Tags from Calcifying Cells of the
           Marine Coccolithophorid, Emiliana huxleyi
JOURNAL    Unpublished (2003)
COMMENT    Contact: Betsy Read
           Department of Biological Sciences
           California State University San Marcos
           333 S. Twin Oaks Valley Road, San Marcos, CA 92096-0001, USA
           Tel: 760 750 4129
           Email: bread@csusm.edu.

FEATURES   source
            1..4096
               /organism="Emiliana huxleyi"
               /mol_type="mRNA"
               /strain="1516"
               /db_xref="taxon:2903"
               /clone_stage="late log growth phase"
               /gene_lib="Preamplified custom cDNA library in pMAB58
               (ResGen, Invitrogen Inc.)"
               /note="Emiliana huxleyi grown in Phosphate limited media
               (Guillard's F/50 media)."

ORIGIN
Query Match      97.2%; Score 485.8; DB 14; Length 4096;
Best Local Similarity 99.6%; Pred. No. 1.4e-117;
Matches 487; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      12 CAGGTGGCACTTTTCGGGAAATGTCGCGGAACCCCTATTGTTTATTCTTCTAAATAC 71
Db      3510 CAGGTGGCACTTTTCGGGAAATGTCGCGGAACCCCTATTGTTTATTCTTCTAAATAC 3569

QY      72 ATTCAATATGATTCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATAATATTGAA 131
Db      3570 ATTCAATATGATTCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATAATATTGAA 3629

QY      132 AAAGGAGAGATGATGAGTATTCAACATTCGCTGCGCCCTATTTCCTTTTGGCGCAT 191

```

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 614.

Location/Qualifiers

FEATURES

source

1. .614
 /organism="Sterkiella histriomuscorum"
 /mol_type="genomic DNA"
 /db_xref="taxon:94289"
 /clone="UUGCLO0004F10"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /notes="Oxytricha plasmid UUGC10 library"
 /vector="PMD42nv; Purified macronuclear chromosomal DNA from Oxytricha trifallax was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. Vector DNA was prepared from a derivative of pMD42 (GI:4732114|GB|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. Coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 97.4%; Score 487.2; DB 29; Length 614;
 Best Local Similarity 99.2%; Pred. No. 5.7e-118;
 Matches 489; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 ACCTCAGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 67
 DB 88 ACCTCAGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 147

QY 68 ATACATTCATATGATTCGGCTCATGACACATACCCCTGATAATGCTTCAATAAT 127
 DB 148 ATACATTCATATGATTCGGCTCATGACACATACCCCTGATAATGCTTCAATAAT 207

QY 128 TGAAGAAGAGATGATGATATCAACATTCGCTGTCGCCCTATTCCCTTTTGG 187
 DB 208 TGAAGAAGAGATGATGATATCAACATTCGCTGTCGCCCTATTCCCTTTTGG 267

QY 188 GCATTTTGCCTTCTCTGTTTGTCTACCCAGAAACGCTGTAAGATAAAGATGCTGAA 247
 DB 268 GCATTTTGCCTTCTCTGTTTGTCTACCCAGAAACGCTGTAAGATAAAGATGCTGAA 327

QY 248 GATCAGTTGGGTGCAACAGTGGGTATCATCGAACTGATCTCAAGCGGTAGATCCTT 307
 DB 328 GATCAGTTGGGTGCAACAGTGGGTATCATCGAACTGATCTCAAGCGGTAGATCCTT 387

QY 308 GAGAGTTTTCGCCCGGAAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
 DB 388 GAGAGTTTTCGCCCGGAAGAACGTTTTCCTAATGATGAGCACTTTTAAAGTTCTGCTATGT 447

QY 368 GGCGCGGTATATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGGCATACACTAT 427
 DB 448 GGCGCGGTATATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGGCATACACTAT 507

QY 428 TCTCAGATGACTTGGTGGTACTCATCAGTACAGAAAGCATCTTACGGATGCGATG 487
 DB 508 TCTCAGATGACTTGGTGGTACTCATCAGTACAGAAAGCATCTTACGGATGCGATG 567

QY 488 ACAGTAAGAGAT 500
 DB 568 ACAGTAAGAGAT 580

RESULT 2

CD279814

LOCUS

G44222.86 NCI CGAP_Zemb3 Danio rerio cDNA clone IMAGE:6321548 5',

DEFINITION

mRNA sequence.

ACCESSION

CD279814

VERSION

CD279814.1 GI:31057590

KEYWORDS

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE

1. (bases 1 to 616)

AUTHORS

Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A.,
 Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
 Expressed sequence tags from NCI_CGAP_Zemb3, a Danio rerio
 embryonic library
 Unpublished (2003)

TITLE

Contact: Chen F.

JOURNAL

COMMENT

Exelixis, Inc.

170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA

Tel: 650 837 7000

Fax: 650 837 8300

Email: fchen@exelixis.com

DNA Sequencing by: Exelixis, Inc. Clone distribution information

can be found through the I.M.A.G.E. Consortium/LINL at:

http://image.llnl.gov

Plate: 13761 row: K column: 21

High quality sequence stop: 616.

FEATURES

source

1. .616

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7985"

/clone="IMAGE:6321548"

/tissue_type="embryo"

/lab_host="DH10B (T1-resistant)"

/clone_lib="NCI CGAP_Zemb3"

/note="vector: pCMV-Sport6.1; Site1: EcoRV; Site2: NotI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 Kb. Constructed by J. Wang (Research Genetics,
 Invitrogen Corp) from tissue donated by L. Zon (Harvard
 University). Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 97.4%; Score 487; DB 14; Length 616;
 Best Local Similarity 99.0%; Pred. No. 6.4e-118;
 Matches 490; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTACGTGAGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTCT 65
 DB 57 CTATTTAGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTCT 116

QY 66 AATACATTCATATGATTCGGCTCATGACACATTAACCTGATAATGCTTCAATAAT 125
 DB 117 AATACATTCATATGATTCGGCTCATGACACATTAACCTGATAATGCTTCAATAAT 176

QY 126 ATTGAAAAGGAGAGATGATGATTCAACATTTCCGTGCGCCTATTCCCTTTTGTG 185
 DB 177 ATTGAAAAGGAGAGATGATGATTCAACATTTCCGTGCGCCTATTCCCTTTTGTG 236

QY 186 CGCATTTTGCCTTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGATGCTG 245
 DB 237 CGCATTTTGCCTTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGATGCTG 296

QY 246 AAGATCAGTTGGTGCAGAGTGGTTTACATCGAATCGATCTCAACAGCGGTAAGATCC 305
 DB 297 AAGATCAGTTGGTGCAGAGTGGTTTACATCGAATCGATCTCAACAGCGGTAAGATCC 356

QY 306 TTGAGAGTTTTCGCCCGGAAGAACGTTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTAT 365
 DB 357 TTGAGAGTTTTCGCCCGGAAGAACGTTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTAT 416

QY 366 GTGCGCGGTATTATCCGTCTTGACCGCGGCAAGAGCAACTCGGTCCGCGCATACACT 425
 DB 417 GTGCGCGGTATTATCCGTCTTGACCGCGGCAAGAGCAACTCGGTCCGCGCATACACT 476

QY 426 ATTCTCAGATGACTTGGTTGAGTACTCACAGTCAAGAAAGCATCTTACGGATGCGCA 485

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:25:56 ; Search time 3054.17 Seconds
(without alignments)
4888.751 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500
Perfect score: 500
Sequence: 1 gttactacgtcaggtggca.....tggcatgacagtaagaagaat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rtd:*
 - 26: em_gss_phg:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	487.2	97.4	614	29	CC818400 100004F10
2	487	97.4	616	14	CD279814 G44222.86
3	485.8	97.2	860	13	BQ817108 1030061G0
4	485.8	97.2	4096	14	CF753623 EST-Cont1

ALIGNMENTS

RESULT 1	CC818400	614 bp	DNA	linear	GSS 17-JUL-2003
LOCUS	100004F10R	Oxytricha plasmid	UUGC10 library	Starkieella	
DEFINITION	histriomuscorum	Genomic clone	UUGC100004F10 R	Genomic survey	
sequence.	CC818400				
ACCESSION	CC818400				
VERSION	CC818400.1	GI:32897704			
KEYWORDS	GSS.				
SOURCE	Sterkiella histriomuscorum				
ORGANISM	Sterkiella histriomuscorum (Oxytricha trifallax)				
REFERENCE	Dunn,D., Doak,T., Herrick,G. and Weiss,R.				
AUTHORS	Paired end reads from plasmid inserts of Oxytricha trifallax				
TITLE	macronuclear chromosomes				
JOURNAL	Unpublished (2003)				
COMMENT	Contact: Robert B. Weiss				
	University of Utah Genome Center				
	University of Utah				
	Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT				
	84112, USA				
	Tel: 801 585 5606				
	Fax: 801 585 7177				
	Email: dunn@genetics.utah.edu				
	Plate: 0004 row: F column: 10				

Db 7708 GCGCGGTATTATCCCGTTGACCGCGGCAAGCAACTCGGTGCGCGCATACACTAT 7649
 QY 428 TCTCAGATGACTTGGTTGAGTACTCACCAGTCA CAGAAAAGCATCTTTACGGATGGCATG 487
 Db 7648 TCTCAGATGACTTGGTTGAGTACTCACCAGTCA CAGAAAAGCATCTTTACGGATGGCATG 7589
 QY 488 ACAGTAAGAGAAT 500
 Db 7588 ACAGTAAGAGAAT 7576

Search completed: April 22, 2004, 13:06:14
 Job time : 340.946 secs

QY 248 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 307
Db 2808 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 2749
QY 308 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 2748 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 2689
QY 368 GGGCGGGTATATCCCGTGTGTAGCGCGGCGAAGCAACTCGGTGCGCGCATACACTAT 427
Db 2688 GGGCGGGTATATCCCGTGTGTAGCGCGGCGAAGCAACTCGGTGCGCGCATACACTAT 2629
QY 428 TCTCAGATGACTTGGTTGAGTACTCACCAGTCCACAGAAAGCACTTTACGGATGGCATG 487
Db 2628 TCTCAGATGACTTGGTTGAGTACTCACCAGTCCACAGAAAGCACTTTACGGATGGCATG 487
QY 488 ACAGTAAGAGAAAT 500
Db 2568 ACAGTAAGAGAAAT 2556
RESULT 14
US-10-491-121-12/c
; Sequence 12, Application US/10491121
; GENERAL INFORMATION:
; APPLICANT: NABEL, GARY
; APPLICANT: YANG, ZHI-YONG
; APPLICANT: SULLIVAN, NANCY
; APPLICANT: SANCHEZ, ANTHONY
; TITLE OF INVENTION: Development of a Preventive Vaccine for
; FILE REFERENCE: NIH221.001NP
; CURRENT APPLICATION NUMBER: US/10/491,121
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: PCT/US02/30251
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/326476
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 8131
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct padapt Ebola GP(R) (dTM)
US-10-491-121-12
Query Match 98.3%; Score 491.4; DB 6; Length 8131;
Best Local Similarity 99.8%; Pred. No. 3.2e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACCTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTCTAA 67
Db 8053 ACCTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTCTAA 7994
QY 68 ATACATTCAAATATGATCGGTCTATGAGACAATAACCCCTGATAAATGCTTCAATAAT 127
Db 7993 ATACATTCAAATATGATCGGTCTATGAGACAATAACCCCTGATAAATGCTTCAATAAT 7934
QY 128 TGAAGAAGGAGATGAGTATCAACATTTCCGTGTCGCCCTATTCCCTTTTTCGG 187
Db 7933 TGAAGAAGGAGATGAGTATCAACATTTCCGTGTCGCCCTATTCCCTTTTTCGG 7874
QY 188 GCATTTTGCCTTCTCTTTTCTCACCAGAAACGCTGTAAGTAAGATGCTGAA 247
Db 7873 GCATTTTGCCTTCTCTTTTCTCACCAGAAACGCTGTAAGTAAGATGCTGAA 7814
QY 248 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 307
Db 7813 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 7754
QY 308 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367

Db 7753 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 7694
QY 368 GGGCGGGTATATCCCGTGTGTAGCGCGGCGAAGCAACTCGGTGCGCGCATACACTAT 427
Db 7693 GGGCGGGTATATCCCGTGTGTAGCGCGGCGAAGCAACTCGGTGCGCGCATACACTAT 7634
QY 428 TCTCAGATGACTTGGTTGAGTACTCACCAGTCCACAGAAAGCACTTTACGGATGGCATG 487
Db 7633 TCTCAGATGACTTGGTTGAGTACTCACCAGTCCACAGAAAGCACTTTACGGATGGCATG 7574
QY 488 ACAGTAAGAGAAAT 500
Db 7573 ACAGTAAGAGAAAT 7561
RESULT 15
US-10-491-121-20/c
; Sequence 20, Application US/10491121
; GENERAL INFORMATION:
; APPLICANT: NABEL, GARY
; APPLICANT: YANG, ZHI-YONG
; APPLICANT: SULLIVAN, NANCY
; APPLICANT: SANCHEZ, ANTHONY
; TITLE OF INVENTION: Development of a Preventive Vaccine for
; FILE REFERENCE: NIH221.001NP
; CURRENT APPLICATION NUMBER: US/10/491,121
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: PCT/US02/30251
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/326476
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 8146
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct padapt EbolaGP (IC) (dTM)
US-10-491-121-20
Query Match 98.3%; Score 491.4; DB 6; Length 8146;
Best Local Similarity 99.8%; Pred. No. 3.2e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACCTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTCTAA 67
Db 8068 ACCTCAGGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTTCTAA 8009
QY 68 ATACATTCAAATATGATCGGTCTATGAGACAATAACCCCTGATAAATGCTTCAATAAT 127
Db 8008 ATACATTCAAATATGATCGGTCTATGAGACAATAACCCCTGATAAATGCTTCAATAAT 7949
QY 128 TGAAGAAGGAGATGAGTATCAACATTTCCGTGTCGCCCTATTCCCTTTTTCGG 187
Db 7948 TGAAGAAGGAGATGAGTATCAACATTTCCGTGTCGCCCTATTCCCTTTTTCGG 7889
QY 188 GCATTTTGCCTTCTCTTTTCTCACCAGAAACGCTGTAAGTAAGATGCTGAA 247
Db 7888 GCATTTTGCCTTCTCTTTTCTCACCAGAAACGCTGTAAGTAAGATGCTGAA 7829
QY 248 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 307
Db 7828 GATCAGTTGGTGCACGAGTGGGTTACATCGAATCTCAACAGCGGTAAGATCCTT 7769
QY 308 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 7768 GAGAGTTTTCGCCGCCGGAAGACCTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT 7709
QY 368 GGGCGGGTATATCCCGTGTGTAGCGCGGCGAAGCAACTCGGTGCGCGCATACACTAT 427


```

Db 3091 TCTCAGATGACTGGTTGAGTACTCACCAGTACAGAAAGCATCTTACGGATGGCATG 3150
QY 488 ACAGTAAGAGAAAT 500
Db 3151 ACAGTAAGAGAAAT 3163

RESULT 12
US-10-491-121-35/c
; Sequence 35, Application US/10491121
; GENERAL INFORMATION:
; APPLICANT: NABEL, GARY
; APPLICANT: YANG, ZHI-YONG
; APPLICANT: SULLIVAN, NANCY
; APPLICANT: SANCHEZ, ANTHONY
; TITLE OF INVENTION: Development of a Preventive Vaccine for
; TITLE OF INVENTION: Filovirus Infection in Primates
; FILE REFERENCE: NIH221.001NP
; CURRENT APPLICATION NUMBER: US/10/491.121
; CURRENT FILING DATE: 2004-03-25
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/326476
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 7711
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct pAdapt Lassa GP
US-10-491-121-35

Query Match 98.3%; Score 491.4; DB 6; Length 7711;
Best Local Similarity 99.8%; Pred. No. 3.1e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACGTCAGTGGCCTTTTCGGGGAATGTCGCGGAACCCCTATTTCGTTATTTTCTAA 67
Db 7633 ACGTCAGTGGCCTTTTCGGGGAATGTCGCGGAACCCCTATTTCGTTATTTTCTAA 7574
QY 68 ATACATTCATATATGATCCGCTCATGAGCAATTAACCCCTGATTAATGCTTCAATAAT 127
Db 7573 ATACATTCATATATGATCCGCTCATGAGCAATTAACCCCTGATTAATGCTTCAATAAT 7514
QY 128 TGAAGAGAGAGATGATGAGTATTCACATTCGCTGTCGCCCTTATTCCTCTTTTGG 187
Db 7513 TGAAGAGAGAGATGATGAGTATTCACATTCGCTGTCGCCCTTATTCCTCTTTTGG 7454
QY 188 GCATTTTGCCTTCCTCTTTTTCGTCACCCAGAACGCTGTGAAAGTAAAGATGCTGAA 247
Db 7453 GCATTTTGCCTTCCTCTTTTTCGTCACCCAGAACGCTGTGAAAGTAAAGATGCTGAA 7394
QY 248 GATCAGTTGGGTGACAGAGTGGTTACATCGAATGATCTCAACACGCTAGATCCCT 307
Db 7393 GATCAGTTGGGTGACAGAGTGGTTACATCGAATGATCTCAACACGCTAGATCCCT 7334
QY 308 GAGAGTTTCCCGGAGAAAGCTTCTCCATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 7333 GAGAGTTTCCCGGAGAAAGCTTCTCCATGATGAGCACTTTTAAAGTTCTGCTATGT 7274
QY 368 GGCGCGGTATTCCTCGTGTGACGCGGCAAGCAACTCGTGCGCGCATACACTAT 427
Db 7273 GGCGCGGTATTCCTCGTGTGACGCGGCAAGCAACTCGTGCGCGCATACACTAT 7214
QY 428 TCTCAGATGACTGGTTGAGTACTCAACAGTACAGAAAGCATCTTACGGATGGCATG 487
Db 7213 TCTCAGATGACTGGTTGAGTACTCAACAGTACAGAAAGCATCTTACGGATGGCATG 7154
QY 488 ACAGTAAGAGAAAT 500

```

```

Db 7153 ACAGTAAGAGAAAT 7141

RESULT 13
US-10-313-392-20/c
; Sequence 20, Application US/10313392
; GENERAL INFORMATION:
; APPLICANT: Joliffe, Linda K.
; APPLICANT: Zivlin, Robert A.
; APPLICANT: Pulito, Virginia L.
; TITLE OF INVENTION: CDR-GRAFTED ANTI-TISSUE FACTOR
; TITLE OF INVENTION: ANTIBODIES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/313,392
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: 530
; APPLICATION NUMBER: US/08/480,120
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9598
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..711
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-313-392-20

Query Match 98.3%; Score 491.4; DB 6; Length 7864;
Best Local Similarity 99.8%; Pred. No. 3.2e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACGTCAGTGGCCTTTTCGGGGAATGTCGCGGAACCCCTATTTCGTTATTTTCTAA 67
Db 3048 ACGTCAGTGGCCTTTTCGGGGAATGTCGCGGAACCCCTATTTCGTTATTTTCTAA 2989
QY 58 ATACATTCATATATGATCCGCTCATGAGCAATTAACCCCTGATTAATGCTTCAATAAT 127
Db 2988 ATACATTCATATATGATCCGCTCATGAGCAATTAACCCCTGATTAATGCTTCAATAAT 2929
QY 128 TGAAGAGAGAGATGATGAGTATTCACATTCGCTGTCGCCCTTATTCCTCTTTTGG 187
Db 2928 TGAAGAGAGAGATGATGAGTATTCACATTCGCTGTCGCCCTTATTCCTCTTTTGG 2869
QY 188 GCATTTTGCCTTCCTCTTTTTCGTCACCCAGAAACGCTGTGAAAGTAAAGATGCTGAA 247
Db 2868 GCATTTTGCCTTCCTCTTTTTCGTCACCCAGAAACGCTGTGAAAGTAAAGATGCTGAA 2809

```

Db 4409 GATCAGTTGGTGCACGAGTGGGTTACATCGAAGTCTCAACAGCGGTAAAGTCCTT 4350
QY 308 GAGAGTTTTCGCCCGGAGAACGTTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 4349 GAGAGTTTTCGCCCGGAGAACGTTTTCATGATGAGCACTTTTAAAGTTCTGCTATGT 4290
QY 368 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 427
Db 4289 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 4230
QY 428 TCTCAGAAATGACGTTGGTGGTACTCACCAGTACAGAAAAGCACTTTACGGATGGCATG 487
Db 4229 TCTCAGAAATGACGTTGGTGGTACTCACCAGTACAGAAAAGCACTTTACGGATGGCATG 4170
QY 488 ACAGTAAGAGAAT 500
Db 4169 ACAGTAAGAGAAT 4157

RESULT 10
US-10-491-121-36/c
; Sequence 36, Application US/10491121
; GENERAL INFORMATION:
; APPLICANT: NABEL, GARY
; APPLICANT: YANG, ZHI-YONG
; APPLICANT: SULLIVAN, NANCY
; APPLICANT: SANCHEZ, ANTHONY
; TITLE OF INVENTION: Development of a Preventive Vaccine for
; TITLE OF INVENTION: Filovirus Infection in Primates
; FILE REFERENCE: NIH221.001NP
; CURRENT APPLICATION NUMBER: US/10/491,121
; CURRENT FILING DATE: 2004-03-26
; PRIOR FILING DATE: 2002-09-24
; PRIOR FILING DATE: 2002-09-24
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 7522
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct pAdapt Lassa GP(dTM)
US-10-491-121-36

Query Match 98.3%; Score 491.4; DB 6; Length 7522;
Best Local Similarity 99.8%; Pred. No. 3.1e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAA 67
Db 7444 ACGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAA 7385
QY 68 ATACATTCAAATATGATATCGGCTCATGAGACAAATACCCCTGATAAATGCTTCAATAAT 127
Db 7384 ATACATTCAAATATGATATCGGCTCATGAGACAAATACCCCTGATAAATGCTTCAATAAT 7325
QY 128 TGAAGAAGAGAGATGATGATTCACCAATTCGCTGCGCCCTATTCCCTTTTTCG 187
Db 7324 TGAAGAAGAGAGATGATGATTCACCAATTCGCTGCGCCCTATTCCCTTTTTCG 7265
QY 188 GCATTTTGCCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGATGCTGAA 247
Db 7264 GCATTTTGCCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGATGCTGAA 7205
QY 248 GATCAGTTGGGTGCAGAGTGGTGTACATCGAATCGGATCTCAACAGCGGTAAAGTCTTT 307
Db 7204 GATCAGTTGGGTGCAGAGTGGTGTACATCGAATCGGATCTCAACAGCGGTAAAGTCTTT 7145
QY 308 GAGAGTTTTCGCCCGGAGAACGTTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 7144 GAGAGTTTTCGCCCGGAGAACGTTTTCATGATGAGCACTTTTAAAGTTCTGCTATGT 7085

QY 368 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 427
Db 7084 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 7025
QY 428 TCTCAGAAATGACGTTGGTGGTACTCACCAGTACAGAAAAGCACTTTACGGATGGCATG 487
Db 7024 TCTCAGAAATGACGTTGGTGGTACTCACCAGTACAGAAAAGCACTTTACGGATGGCATG 6965
QY 488 ACAGTAAGAGAAT 500
Db 6964 ACAGTAAGAGAAT 6952

RESULT 11
US-10-612-410-1
; Sequence 1, Application US/10612410
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; TITLE OF INVENTION: Methods and Compositions for the Production, Identification and
; TITLE OF INVENTION: Purification of Fusion Proteins
; FILE REFERENCE: 0942.5510003
; CURRENT APPLICATION NUMBER: US/10/612,410
; CURRENT FILING DATE: 2003-07-03
; PRIOR FILING DATE: 2003-07-03
; PRIOR FILING DATE: 2002-07-08
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 7618
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: pET104-DEST
US-10-612-410-1

Query Match 98.3%; Score 491.4; DB 6; Length 7618;
Best Local Similarity 99.8%; Pred. No. 3.1e-136;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAA 67
Db 2671 ACGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAA 2730
QY 68 ATACATTCAAATATGATATCGGCTCATGAGACAAATACCCCTGATAAATGCTTCAATAAT 127
Db 2731 ATACATTCAAATATGATATCGGCTCATGAGACAAATACCCCTGATAAATGCTTCAATAAT 2790
QY 128 TGAAGAAGAGAGATGATGATTCACCAATTCGCTGCGCCCTATTCCCTTTTTCG 187
Db 2791 TGAAGAAGAGAGATGATGATTCACCAATTCGCTGCGCCCTATTCCCTTTTTCG 2850
QY 188 GCATTTTGCCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGATGCTGAA 247
Db 2851 GCATTTTGCCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGATGCTGAA 2910
QY 248 GATCAGTTGGGTGCAGAGTGGTGTACATCGAATCGGATCTCAACAGCGGTAAAGTCTTT 307
Db 2911 GATCAGTTGGGTGCAGAGTGGTGTACATCGAATCGGATCTCAACAGCGGTAAAGTCTTT 2970
QY 308 GAGAGTTTTCGCCCGGAGAACGTTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 2971 GAGAGTTTTCGCCCGGAGAACGTTTTCATGATGAGCACTTTTAAAGTTCTGCTATGT 3030
QY 368 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 427
Db 3031 GGGCGCGTATTATCCCGTGTGACGCGGCGAAGACCACTCGGTGCGCGCATACACTAT 3090
QY 428 TCTCAGAAATGACGTTGGTGGTACTCACCAGTACAGAAAAGCACTTTACGGATGGCATG 487

```

; APPLICANT: SENALDI, GIORGIO
; TITLE OF INVENTION: COMBINATION THERAPY FOR CONDITIONS LEADING TO BONE LOSS
; FILE REFERENCE: A-378CIP5C
; CURRENT APPLICATION NUMBER: US/09/613,591F
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 09/457,647
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: US 09/350,670
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: US 08/706,945
; PRIOR FILING DATE: 1996-09-03
; PRIOR APPLICATION NUMBER: US 08/577,788
; PRIOR FILING DATE: 1995-12-22
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 176
; LENGTH: 6037
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4760)..(6025)
US-09-613-591F-176

```

Query Match	98.3%;	Score 491.4;	DB 5;	Length 6037;
Best Local Similarity	99.8%;	Pred. No. 2.9e-136;		
Matches 492;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	8	ACGTGAGTGGCACATTTTCGGGGAAATGTCGGCGGAACCCCTATTGTGTTATTATTTTCTAA	67	
Db	3876	ACGTGAGTGGCACATTTTCGGGGAAATGTCGGCGGAACCCCTATTGTGTTATTATTTTCTAA	3817	
Qy	68	ATACATTCAAAATATGATCCGCTCATGAGACAAATAACCCCTGATAAATGCTTCAATAATAT	127	
Db	3816	ATACATTCAAAATATGATCCGCTCATGAGACATATACCCCTGATTAATGCTTCAATAATAT	3757	
Qy	128	TGAAAAAGAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCTTTTTCG	187	
Db	3756	TGAAAAAGAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCTTTTTCG	3697	
Qy	188	GCATTTTCGCCCTTCGTGTTTGTCCACCCAGAACGCTGGTGAAGTAAAGATGCTGAA	247	
Db	3696	GCATTTTCGCCCTTCGTGTTTGTCCACCCAGAACGCTGGTGAAGTAAAGATGCTGAA	3637	
Qy	248	GATCAGTTGGGTGCACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATCCCT	307	
Db	3636	GATCAGTTGGGTGCACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATCCCT	3577	
Qy	308	GAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCGTATGT	367	
Db	3576	GAGAGTTTTCGCCCGGAGAACGTTTCCAAATGATGAGCACTTTTAAAGTTCGTATGT	3517	
Qy	368	GGCGCGGTATTATCCCGTGTGACCGCGGCAAGACAACTCGGTCCGCGATACACTAT	427	
Db	3516	GGCGCGGTATTATCCCGTGTGACCGCGGCAAGACAACTCGGTCCGCGATACACTAT	3457	
Qy	428	TCTCAGATGACTGGTTGAGTACTCACACAGTACACAGAAAAGCATCTTACGGATGGCATG	487	
Db	3456	TCTCAGATGACTGGTTGAGTACTCACACAGTACACAGAAAAGCATCTTACGGATGGCATG	3397	
Qy	488	ACAGTAAGAGAAAT	500	
Db	3396	ACAGTAAGAGAAAT	3384	

```

RESULT 9
US-10-313-392-15/c
; Sequence 15, Application US/10313392
; GENERAL INFORMATION:
; APPLICANT: Joliffe, Linda K.
; Zivin, Robert A.
; Pulito, Virginia L
; TITLE OF INVENTION: CDR-GRAFT
;

```

ANTIBODIES AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: United States

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/313,392

FILING DATE: 07-Jun-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/480,120

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: DiGiglio, Frank S.

REGISTRATION NUMBER: 31,346

REFERENCE/DOCKET NUMBER: 9598

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 742-4343

TELEFAX: (516) 742-4366

TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 7073 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: CDS

LOCATION: 61..717

FEATURE:

NAME/KEY: CDS

LOCATION: 1111..1146

FEATURE:

NAME/KEY: CDS

LOCATION: 1268..1594

FEATURE:

NAME/KEY: CDS

LOCATION: 1692..2012

SEQUENCE DESCRIPTION: SEQ ID NO: 15:

US-10-313-392-15

Query Match	98.3%	Score	491.4	DB	6	Length	7073
Best Local Similarity	99.8%	Pred.	No. 3e-136				
Matches	492	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
QY	8	ACGTCAGGTGGCACATTTTCGGGGAATGTGCGCGGAACCCCTATTGTGTTATTTTCTAA	67				
Db	4649	ACGTCAGGTGGCACATTTTCGGGGAATGTGCGCGGAACCCCTATTGTGTTATTTTCTAA	4590				
QY	68	ATACATTTCAAATATGATCCGTCATGAGACAAATACCCGTGATAAATGCTCAATAAATAT	127				
Db	4589	ATACATTTCAAATATGATCCGTCATGAGACAAATACCCGTGATAAATGCTCAATAAATAT	4530				
QY	128	TGAAAAAGGAAGATGATGATATCAACATTTCCGTGTGCGCCTTATATCCCTCTTTTGGG	187				
Db	4529	TGAAAAAGGAAGATGATGATATCAACATTTCCGTGTGCGCCTTATATCCCTCTTTTGGG	4470				
QY	188	GCATTTTGCTTTCCGTGTTTTCCTCACCACAGAAACGCTGGTGAAAGTAAAGATGCTGAA	247				
Db	4469	GCATTTTGCTTTCCGTGTTTTCCTCACCACAGAAACGCTGGTGAAAGTAAAGATGCTGAA	4410				
QY	248	GATCAGTTCGGGTGCACGAGTGGGTATCATCTGAAACCTGGATCTCAAACAGCGGTAAAGATCCCTT	307				

RESULT 6

US-10-467-781A-3/c

; Sequence 3, Application US/10467781A

; GENERAL INFORMATION:

; APPLICANT: Tisi, Laurence C

; TITLE OF INVENTION: Method for detecting DNA polymerisation

; FILE REFERENCE: 620-265

; CURRENT APPLICATION NUMBER: US/10/467,781A

; CURRENT FILING DATE: 2003-08-13

; PRIOR APPLICATION NUMBER: PCT/GB02/00648

; PRIOR FILING DATE: 2002-02-13

; PRIOR APPLICATION NUMBER: GB 0103622.7

; PRIOR FILING DATE: 2001-02-14

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 3

; LENGTH: 5777

; TYPE: DNA

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Plasmid

; NAME/KEY: misc_feature

; LOCATION: 1, 4, 20, 21, 25, 29, 32, 36, 39

; OTHER INFORMATION: n is uncertain.

US-10-467-781A-3

Query Match

Best Local Similarity 98.3%; Score 491.4; DB 6; Length 5777;

Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACCTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAA 67

Db 5412 ACCTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAA 5353

QY 68 ATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAATCTTCAATAATAT 127

Db 5352 ATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAATCTTCAATAATAT 5293

QY 128 TGAAGAAGGAAGATAGTATTCACAAATTTCCGTCGCGCCCTATTCCCTTTTTCGC 187

Db 5292 TGAAGAAGGAAGATAGTATTCACAAATTTCCGTCGCGCCCTATTCCCTTTTTCGC 5233

QY 188 GCATTTTGCCTTCCTCTGTTTTGCTCACCAGAAACGCTGGTGAAGTAAAGATGCTGAA 247

Db 5232 GCATTTTGCCTTCCTCTGTTTTGCTCACCAGAAACGCTGGTGAAGTAAAGATGCTGAA 5173

QY 248 GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 307

Db 5172 GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 5113

QY 308 GAGAGTTTTCGCGCGGAAAGAGTTCTCAATGATGAGACATTTTAAAGTTCTGCTATGT 367

Db 5112 GAGAGTTTTCGCGCGGAAAGAGTTCTCAATGATGAGACATTTTAAAGTTCTGCTATGT 5053

QY 368 GCGCGGTATTATCCCGTGTTCAGCGCGGCGCAAGAGCAACTCGGTGCGCGCATACACTAT 427

Db 5052 GCGCGGTATTATCCCGTGTTCAGCGCGGCGCAAGAGCAACTCGGTGCGCGCATACACTAT 4993

QY 428 TCTCAGAAATGACTTGGTTGAGTACTCACCAGTACAGAAAGCAATCTTACGGATGGCATG 487

Db 4992 TCTCAGAAATGACTTGGTTGAGTACTCACCAGTACAGAAAGCAATCTTACGGATGGCATG 4933

QY 488 ACAGTAAGAGAAAT 500

Db 4932 ACAGTAAGAGAAAT 4920

RESULT 7

US-10-612-410-2

; Sequence 2, Application US/10612410

; GENERAL INFORMATION:

; APPLICANT: Bennett, Robert P.

; TITLE OF INVENTION: Methods and Compositions for the Production, Identification and

; FILE REFERENCE: 0942.5510003

; CURRENT APPLICATION NUMBER: US/10/612,410

; CURRENT FILING DATE: 2003-07-03

; PRIOR APPLICATION NUMBER: 60/393,756

; PRIOR FILING DATE: 2002-07-08

; PRIOR APPLICATION NUMBER: 60/396,627

; PRIOR FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 60/417,172

; PRIOR FILING DATE: 2002-10-10

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 5934

; TYPE: DNA

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: pET104/D-TOPO

US-10-612-410-2

Query Match

Best Local Similarity 98.3%; Score 491.4; DB 6; Length 5934;

Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACCTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAA 67

Db 987 ACCTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATTTCTAA 1046

QY 68 ATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAATCTTCAATAATAT 127

Db 1047 ATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAATCTTCAATAATAT 1106

QY 128 TGAAGAAGGAAGATAGTATTCACAAATTTCCGTCGCGCCCTATTCCCTTTTTCGC 187

Db 1107 TGAAGAAGGAAGATAGTATTCACAAATTTCCGTCGCGCCCTATTCCCTTTTTCGC 1166

QY 188 GCATTTTGCCTTCCTCTGTTTTGCTCACCAGAAACGCTGGTGAAGTAAAGATGCTGAA 247

Db 1167 GCATTTTGCCTTCCTCTGTTTTGCTCACCAGAAACGCTGGTGAAGTAAAGATGCTGAA 1226

QY 248 GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 307

Db 1227 GATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 1286

QY 308 GAGAGTTTTCGCGCGGAAAGAGTTCTCAATGATGAGACATTTTAAAGTTCTGCTATGT 367

Db 1287 GAGAGTTTTCGCGCGGAAAGAGTTCTCAATGATGAGACATTTTAAAGTTCTGCTATGT 1346

QY 368 GCGCGGTATTATCCCGTGTTCAGCGCGGCGCAAGAGCAACTCGGTGCGCGCATACACTAT 427

Db 1347 GCGCGGTATTATCCCGTGTTCAGCGCGGCGCAAGAGCAACTCGGTGCGCGCATACACTAT 1406

QY 428 TCTCAGAAATGACTTGGTTGAGTACTCACCAGTACAGAAAGCAATCTTACGGATGGCATG 487

Db 1407 TCTCAGAAATGACTTGGTTGAGTACTCACCAGTACAGAAAGCAATCTTACGGATGGCATG 1466

QY 488 ACAGTAAGAGAAAT 500

Db 1467 ACAGTAAGAGAAAT 1479

RESULT 8

US-09-613-591F-176/c

; Sequence 176, Application US/09613591F

; GENERAL INFORMATION:

; APPLICANT: BOYLE, WILLIAM J.

; APPLICANT: LACEY, DAVID LEE

; APPLICANT: CALZONE, FRANK J.

; APPLICANT: CHANG, MING-SHI

QY 301 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
QY 361 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 420
Db 361 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 420
QY 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 480
Db 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAGAAAT 500
Db 481 TGGCATGACAGTAAGAGAAAT 500

RESULT 4

PCT-US03-31941A-3
; Sequence 3, Application PC/TUS0331941A
; GENERAL INFORMATION:
; APPLICANT: Syngenta Participations AG
; APPLICANT: Boudreau, Eric
; APPLICANT: Gu, Weining
; APPLICANT: De Framond, Anic
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: Plasmid Transformation
; FILE REFERENCE: 70149WOPCT
; CURRENT APPLICATION NUMBER: PCT/US03/31941A
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/418596
; PRIOR FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 10453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pNY2C plasmid transformation vector (Example 4 II).
PCT-US03-31941A-3

Query Match 100.0%; Score 500; DB 1; Length 10453;
Best Local Similarity 100.0%; Pred. No. 9.6e-139;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACACGTCAGGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 60
Db 7953 GTTAACACGTCAGGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 8012
QY 61 TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 8013 TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 8072
QY 121 ATAATATTGAAAAGAGAGATGAGTATTCAACATTTCCGTGCGCCCTTATTCCTT 180
Db 8073 ATAATATTGAAAAGAGAGATGAGTATTCAACATTTCCGTGCGCCCTTATTCCTT 8132
QY 181 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 8133 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 8192
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGTATTCAACATTTCCGTGCGCCCTTATTCCTT 300
Db 8073 ATAATATTGAAAAGAGAGATGAGTATTCAACATTTCCGTGCGCCCTTATTCCTT 8132
QY 181 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 8133 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 8192
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGTATTCAACATTTCCGTGCGCCCTTATTCCTT 300
Db 8193 TGCTGAAGATCAGTTGGGTGACGAGTGGTATTCAACATTTCCGTGCGCCCTTATTCCTT 8252
QY 301 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 8253 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 8312
QY 361 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 420
Db 8313 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 8372

QY 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 480
Db 8373 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 8432
QY 481 TGGCATGACAGTAAGAGAAAT 500
Db 8433 TGGCATGACAGTAAGAGAAAT 8452

RESULT 5

US-10-680-824A-3
; Sequence 3, Application US/10680824A
; GENERAL INFORMATION:
; APPLICANT: Boudreau, Eric
; APPLICANT: Gu, Weining
; APPLICANT: De Framond, Anic
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: Plasmid Transformation
; FILE REFERENCE: 70149USNP
; CURRENT APPLICATION NUMBER: US/10/680,824A
; CURRENT FILING DATE: 2003-10-07
; PRIOR APPLICATION NUMBER: 60/418596
; PRIOR FILING DATE: 2002-07-10
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 10453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pNY2C plasmid transformation vector (Example 4 II).
US-10-680-824A-3

Query Match 100.0%; Score 500; DB 6; Length 10453;
Best Local Similarity 100.0%; Pred. No. 9.6e-139;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACACGTCAGGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 60
Db 7953 GTTAACACGTCAGGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 8012
QY 61 TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 8013 TTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 8072
QY 121 ATAATATTGAAAAGAGAGATGAGTATTCAACATTTCCGTGCGCCCTTATTCCTT 180
Db 8073 ATAATATTGAAAAGAGAGATGAGTATTCAACATTTCCGTGCGCCCTTATTCCTT 8132
QY 181 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 8133 TTTTCGGCATTTTGCCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 8192
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGTATTCAACATTTCCGTGCGCCCTTATTCCTT 300
Db 8193 TGCTGAAGATCAGTTGGGTGACGAGTGGTATTCAACATTTCCGTGCGCCCTTATTCCTT 8252
QY 301 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 8253 GATCCTTCAGAGTTTTCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 8312
QY 361 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 420
Db 8313 GCTATGTCGCCGGTATTATCCGCTGTGAGCGCGGCAAGAGCAACTCGGTCCGCCAT 8372
QY 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 480
Db 8373 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTACACAGAAAAGCATCTTACGGA 8432
QY 481 TGGCATGACAGTAAGAGAAAT 500
Db 8433 TGGCATGACAGTAAGAGAAAT 8452

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:32:31 ; Search time 338.946 Seconds
(without alignments)
1009.487 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500

Perfect score: 500

Sequence: 1 gttactactcagtcaggca.....tggcatgacagtaagaat 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1381230 seqs, 342161569 residues

Total number of hits satisfying chosen parameters: 2762460

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA New:*
1: /cgn2_6/prodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/prodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/prodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/prodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/prodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/prodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/prodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	500	100.0	3476	6 US-10-296-085A-1	Sequence 1, Appli
C 2	500	100.0	5759	5 US-09-954-483B-14	Sequence 14, Appl
C 3	500	100.0	6148	5 US-09-954-483B-13	Sequence 13, Appl
C 4	500	100.0	10453	1 PCT-US03-31941A-3	Sequence 3, Appli
C 5	500	100.0	10453	6 US-10-680-824A-3	Sequence 3, Appli
C 6	491.4	98.3	5777	6 US-10-467-781A-3	Sequence 3, Appli
C 7	491.4	98.3	5934	6 US-10-612-410-2	Sequence 2, Appli
C 8	491.4	98.3	6037	5 US-09-613-591F-176	Sequence 176, App
C 9	491.4	98.3	7073	6 US-10-313-392-15	Sequence 15, Appl
C 10	491.4	98.3	7522	6 US-10-491-121-36	Sequence 36, Appli
C 11	491.4	98.3	7618	6 US-10-612-410-1	Sequence 1, Appli
C 12	491.4	98.3	7711	6 US-10-491-121-35	Sequence 35, Appl
C 13	491.4	98.3	7864	6 US-10-313-392-20	Sequence 20, Appl
C 14	491.4	98.3	8131	6 US-10-491-121-12	Sequence 12, Appl
C 15	491.4	98.3	8146	6 US-10-491-121-20	Sequence 20, Appl
C 16	491.4	98.3	8199	6 US-10-491-121-29	Sequence 29, Appl
C 17	491.4	98.3	8221	6 US-10-491-121-27	Sequence 27, Appl
C 18	491.4	98.3	8256	6 US-10-491-121-32	Sequence 32, Appl
C 19	491.4	98.3	8338	6 US-10-491-121-26	Sequence 26, Appl
C 20	491.4	98.3	8439	6 US-10-491-121-28	Sequence 28, Appl
C 21	491.4	98.3	10783	6 US-10-491-121-25	Sequence 25, Appl
C 22	491.4	98.3	11198	6 US-10-776-213-23	Sequence 23, Appl
C 23	491.4	98.3	11201	6 US-10-776-213-25	Sequence 25, Appl
C 24	491.4	98.3	11204	6 US-10-776-213-26	Sequence 26, Appl
C 25	491.4	98.3	11427	6 US-10-776-213-24	Sequence 24, Appl
C 26	491.4	98.3	11600	6 US-10-808-758-48	Sequence 48, Appl

C 27 491.4 98.3 11918 6 US-10-612-224-2
C 28 491.4 98.3 11924 1 PCT-US03-31219-7
C 29 491.4 98.3 11924 6 US-10-678-816-7
C 30 491.4 98.3 12008 6 US-10-776-213-27
C 31 491.4 98.3 12242 1 PCT-US03-31219-6
C 32 491.4 98.3 12242 6 US-10-678-816-6
C 33 491.4 98.3 12844 6 US-10-776-213-19
C 34 491.4 98.3 12850 6 US-10-776-213-22
C 35 491.4 98.3 13073 6 US-10-776-213-20
C 36 491.4 98.3 13227 6 US-10-612-224-3
C 37 491.4 98.3 13278 6 US-10-612-224-3
C 38 491.4 98.3 13654 6 US-10-776-213-28
C 39 491.4 98.3 14683 6 US-10-817-950-1
C 40 491.4 98.3 17135 6 US-10-475-962-55
C 41 491.4 98.3 18386 6 US-10-817-950-2
C 42 491.4 98.3 19040 6 US-10-817-950-3
C 43 491.4 98.3 37474 6 US-10-645-883A-7
C 44 490.4 98.1 4100 6 US-10-240-527A-5
C 45 490.4 98.1 4682 6 US-10-240-527A-3

ALIGNMENTS

RESULT 1

US-10-296-085A-1/c

; Sequence 1, Application US/10296085A

; GENERAL INFORMATION:

; APPLICANT: DAVID M. NEVILLE

; APPLICANT: JERRY T. THOMPSON

; APPLICANT: JUNG-HEE WOO

; APPLICANT: HUAIZHONG HU

; APPLICANT: SHENGLIN MA

; APPLICANT: JONATHAN MARK HEXHAM

; APPLICANT: MARY ELLEN DIGAN

; TITLE OF INVENTION: IMMUNOTOXIN FUSION PROTEINS AND MEANS

; FILE REFERENCE: 14028.029402

; CURRENT APPLICATION NUMBER: US/10/296,085A

; PRIOR FILING DATE: 2002-11-18

; PRIOR APPLICATION NUMBER: 09/573,797

; PRIOR FILING DATE: 2000-05-18

; PRIOR APPLICATION NUMBER: 09/380,484

; PRIOR FILING DATE: 1999-09-03

; PRIOR APPLICATION NUMBER: 09/389,565

; PRIOR FILING DATE: 1999-09-03

; PRIOR APPLICATION NUMBER: PCT/US98/04303

; PRIOR FILING DATE: 1998-03-05

; PRIOR APPLICATION NUMBER: 08/739,703

; PRIOR FILING DATE: 1996-10-29

; PRIOR APPLICATION NUMBER: 60/039,987

; PRIOR FILING DATE: 1997-03-05

; PRIOR APPLICATION NUMBER: 60/008,104

; PRIOR FILING DATE: 1995-10-30

; NUMBER OF SEQ ID NOS: 122

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 3476

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:/Note =

; OTHER INFORMATION: synthetic construct

US-10-296-085A-1

Query Match 100.0%; Score 500; DB 6; Length 3476;
Best Local Similarity 100.0%; Pred. No. 6.5e-139;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTAACTAGTCAGTGGCATTTCGGGAAATGCGCGAATGCGGACCCCTATTGTTATT 60
Db 3476 GTTAACTAGTCAGTGGCATTTCGGGAAATGCGCGAATGCGGACCCCTATTGTTATT 3417

Db 301 GATCTTGAGAGTTTTCGCCCGGAAGACGTTCTCAAATGATGACCACTTTTAAAGTTCT 360
Qy 361 GCTATGTGGCGGGTATTATCCGCTGTGAGCGCCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 361 GCTATGTGGCGGGTATTATCCGCTGTGAGCGCCGGCAAGAGCAACTCGGTGCGCGCAT 420
Qy 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Qy 481 TGGCATGACAGTAAGAGAAT 500
Db 481 TGGCATGACAGTAAGAGAAT 500

RESULT 12
US-09-811-361-1
; Sequence 1, Application US/09811361
; Publication No. US2004040046A1
; GENERAL INFORMATION:
; APPLICANT: Guenther, Catherine
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; FILE REFERENCE: R-125
; CURRENT APPLICATION NUMBER: US/09/811,361
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,348
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage Vector
US-09-811-361-1

Query Match 100.0%; Score 500; DB 13; Length 4768;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTAACACGTCAGTGGCGACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATT 60
Db 1 GTTAACACGTCAGTGGCGACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATT 60

Qy 61 TTCTAAATACATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAATGCTTCA 120
Db 61 TTCTAAATACATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAATGCTTCA 120

Qy 121 ATAATATTGAAAAGGAGATGAGTATTCAACTTCGCTGTCGCCCTTATCCCTT 180
Db 121 ATAATATTGAAAAGGAGATGAGTATTCAACTTCGCTGTCGCCCTTATCCCTT 180

Qy 181 TTTTGGCGCATTTTGCCTTCTCTGTTTGTCTCACCAGAAACGCTGTAAGTAAAGA 240
Db 181 TTTTGGCGCATTTTGCCTTCTCTGTTTGTCTCACCAGAAACGCTGTAAGTAAAGA 240

Qy 241 TGCTGAAGATCAGTTGGGTGCAAGTGGTTATCATGAACTCGATCTCAACAGCGGTAA 300
Db 241 TGCTGAAGATCAGTTGGGTGCAAGTGGTTATCATGAACTCGATCTCAACAGCGGTAA 300

Qy 301 GATCCTTTGAGAGTTTTCGCCCGGGAAGAACGTTCTTCCATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTTGAGAGTTTTCGCCCGGGAAGAACGTTCTTCCATGATGAGCACTTTTAAAGTTCT 360

Qy 361 GCTATGTGGCGGGTATTATCCCGGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 361 GCTATGTGGCGGGTATTATCCCGGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 420

Qy 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480

Qy 481 TGGCATGACAGTAAGAGAAT 500
Db 481 TGGCATGACAGTAAGAGAAT 500

RESULT 13
US-09-815-937-1
; Sequence 1, Application US/09815937
; Publication No. US20020088012A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; FILE REFERENCE: R-611
; CURRENT APPLICATION NUMBER: US/09/815,937
; PRIOR FILING DATE: 2001-03-22
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/221,485
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-937-1

Query Match 100.0%; Score 500; DB 13; Length 4768;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTAACACGTCAGTGGCGACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATT 60
Db 1 GTTAACACGTCAGTGGCGACTTTTCGGGGAATGTCGGGGAACCCCTATTGTTTATT 60

Qy 61 TTCTAAATACATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAATGCTTCA 120
Db 61 TTCTAAATACATTCAAATATGATCGCTCATGAGACAATAACCCCTGATAATGCTTCA 120

Qy 121 ATAATATTGAAAAGGAGATGAGTATTCAACTTCGCTGTCGCCCTTATCCCTT 180
Db 121 ATAATATTGAAAAGGAGATGAGTATTCAACTTCGCTGTCGCCCTTATCCCTT 180

Qy 181 TTTTGGCGCATTTTGCCTTCTCTGTTTGTCTCACCAGAAACGCTGTAAGTAAAGA 240
Db 181 TTTTGGCGCATTTTGCCTTCTCTGTTTGTCTCACCAGAAACGCTGTAAGTAAAGA 240

Qy 241 TGCTGAAGATCAGTTGGGTGCAAGTGGTTATCATGAACTCGATCTCAACAGCGGTAA 300
Db 241 TGCTGAAGATCAGTTGGGTGCAAGTGGTTATCATGAACTCGATCTCAACAGCGGTAA 300

Qy 301 GATCCTTTGAGAGTTTTCGCCCGGGAAGAACGTTCTTCCATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTTGAGAGTTTTCGCCCGGGAAGAACGTTCTTCCATGATGAGCACTTTTAAAGTTCT 360

Qy 361 GCTATGTGGCGGGTATTATCCCGGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 361 GCTATGTGGCGGGTATTATCCCGGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 420

Qy 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACACATTTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480

Qy 481 TGGCATGACAGTAAGAGAAT 500
Db 481 TGGCATGACAGTAAGAGAAT 500

Db 1 GTTAACACGTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
QY 61 TTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 61 TTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
QY 121 ATAATATTGAAAAAGGAAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCCTT 180
Db 121 ATAATATTGAAAAAGGAAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCCTT 180
QY 181 TTCTGCGGCAATTTGCTTCCTGTTTTCGCTCACCAGAAACGCTGTTGGAAGTAAAGA 240
Db 181 TTCTGCGGCAATTTGCTTCCTGTTTTCGCTCACCAGAAACGCTGTTGGAAGTAAAGA 240
QY 241 TCGTGAAGATCAGTTGGGTGACGAGTGGTTCACATTCGAATCGAATCGATCTCAACGCGTAA 300
Db 241 TCGTGAAGATCAGTTGGGTGACGAGTGGTTCACATTCGAATCGAATCGATCTCAACGCGTAA 300
QY 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 360
QY 361 GCTATGTGCGGCGGTATTATCCGCTGTTGACGCGGCAAGAGCAACTCGGTCCGCCAT 420
Db 361 GCTATGTGCGGCGGTATTATCCGCTGTTGACGCGGCAAGAGCAACTCGGTCCGCCAT 420
QY 421 ACACATTTCTCAGAATGACTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACACATTTCTCAGAATGACTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAGAT 500
Db 481 TGGCATGACAGTAAGAGAT 500

RESULT 6

US-10-161-403-121/c
; Sequence 121, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121
; LENGTH: 4223
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pLIT38attBBSRpolyA2 Plasmid
US-10-161-403-121

Query Match 100.0%; Score 500; DB 15; Length 4223;
Best Local Similarity 100.0%; Pred. No. 2e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACACGTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60

Db 3927 GTTAACACGTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 3868
QY 61 TTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 3867 TTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 3808
QY 121 ATAATATTGAAAAAGGAAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCCTT 180
Db 3807 ATAATATTGAAAAAGGAAGAGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCCTT 3748
QY 181 TTCTGCGGCAATTTGCTTCCTGTTTTCGCTCACCAGAAACGCTGTTGGAAGTAAAGA 240
Db 3747 TTCTGCGGCAATTTGCTTCCTGTTTTCGCTCACCAGAAACGCTGTTGGAAGTAAAGA 3688
QY 241 TCGTGAAGATCAGTTGGGTGACGAGTGGTTCACATTCGAATCGAATCGATCTCAACGCGTAA 300
Db 3687 TCGTGAAGATCAGTTGGGTGACGAGTGGTTCACATTCGAATCGAATCGATCTCAACGCGTAA 3628
QY 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 360
Db 3627 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 3568
QY 361 GCTATGTGCGGCGGTATTATCCGCTGTTGACGCGGCAAGAGCAACTCGGTCCGCCAT 420
Db 3567 GCTATGTGCGGCGGTATTATCCGCTGTTGACGCGGCAAGAGCAACTCGGTCCGCCAT 3508
QY 421 ACACATTTCTCAGAATGACTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 3507 ACACATTTCTCAGAATGACTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 3448
QY 481 TGGCATGACAGTAAGAGAT 500
Db 3447 TGGCATGACAGTAAGAGAT 3428

RESULT 7

US-09-816-790-1
; Sequence 1, Application US/09816790
; Patent No. US20020022255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-816-790-1

Query Match 100.0%; Score 500; DB 9; Length 4768;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACACGTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
Db 1 GTTAACACGTCAGGTGGACATTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
QY 61 TTCTAAATACATTCAAATATGATATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120


```
Db 1 GTTAACACGTCAGGTGGACATTTTCGGGGAAATGTGCGCGAAACCCCTATTGTTTATT 60
QY 61 TTCTAAATACATTCAAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 61 TTCTAAATACATTCAAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
QY 121 ATAATATTGAAAAGAGAGATGAGTATTCAGTATTCAAATTTCCGTGTCGCCCTTATTCCCTT 180
Db 121 ATAATATTGAAAAGAGAGATGAGTATTCAGTATTCAAATTTCCGTGTCGCCCTTATTCCCTT 180
QY 181 TTTTGGGCATTTTGCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 181 TTTTGGGCATTTTGCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
QY 241 TGCTGAAGATCAGTTGGGTGACAGTGGGTTTACATCGAATCTGGATCTCAACAGCGGTAA 300
Db 241 TGCTGAAGATCAGTTGGGTGACAGTGGGTTTACATCGAATCTGGATCTCAACAGCGGTAA 300
QY 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
QY 361 GCTATGTGCGCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 361 GCTATGTGCGCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
QY 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAGAAT 500
Db 481 TGGCATGACAGTAAGAGAAT 500
```

RESULT 2

```
US-10-161-403-120
; Sequence 120, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 2847
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pLIT38atcB Plasmid
US-10-161-403-120
```

```
Query Match 100.0%; Score 500; DB 15; Length 2847;
Best Local Similarity 100.0%; Pred. No. 1.6e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTTAACACGTCAGGTGGACATTTTCGGGGAAATGTGCGGGAAACCCCTATTGTTTATT 60
Db 1 GTTAACACGTCAGGTGGACATTTTCGGGGAAATGTGCGGGAAACCCCTATTGTTTATT 60
```

```
QY 61 TTCTAAATACATTCAAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
Db 61 TTCTAAATACATTCAAATATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCA 120
QY 121 ATAATATTGAAAAGAGAGATGAGTATTCAGTATTCAAATTTCCGTGTCGCCCTTATTCCCTT 180
Db 121 ATAATATTGAAAAGAGAGATGAGTATTCAGTATTCAAATTTCCGTGTCGCCCTTATTCCCTT 180
QY 181 TTTTGGGCATTTTGCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 181 TTTTGGGCATTTTGCCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
QY 241 TGCTGAAGATCAGTTGGGTGACAGTGGGTTTACATCGAATCTGGATCTCAACAGCGGTAA 300
Db 241 TGCTGAAGATCAGTTGGGTGACAGTGGGTTTACATCGAATCTGGATCTCAACAGCGGTAA 300
QY 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 301 GATCCTTGAGAGTTTTCGCCCGGAAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
QY 361 GCTATGTGCGCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 361 GCTATGTGCGCGGTATTATCCCGTGTGACGCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
QY 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAGAAT 500
Db 481 TGGCATGACAGTAAGAGAAT 500
```

RESULT 3

```
US-10-161-403-91
; Sequence 91, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10/161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91
; LENGTH: 3438
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pLIT38attBZeo Plasmid
US-10-161-403-91
```

```
Query Match 100.0%; Score 500; DB 15; Length 3438;
Best Local Similarity 100.0%; Pred. No. 1.8e-138;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GTTAACACGTCAGGTGGACATTTTCGGGGAAATGTGCGGGAAACCCCTATTGTTTATT 60
Db 241 GTTAACACGTCAGGTGGACATTTTCGGGGAAATGTGCGGGAAACCCCTATTGTTTATT 300
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:42:41 ; Search time 433.382 Seconds
(without alignments)
5201.676 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500

Perfect score: 500

Sequence: 1 gtaactacgtcaggtggc.....tggcatgacagtaagaagaat 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	2814	15	US-10-161-403-119
2	500	100.0	2847	15	US-10-161-403-120
3	500	100.0	3438	15	US-10-161-403-91
4	500	100.0	3476	15	US-10-435-567-1
5	500	100.0	4223	15	US-10-161-403-111
6	500	100.0	4223	15	US-10-161-403-121
7	500	100.0	4768	9	US-09-816-790-1
8	500	100.0	4768	9	US-09-861-077-1
9	500	100.0	4768	9	US-09-815-825-1
10	500	100.0	4768	9	US-09-815-935-1
11	500	100.0	4768	9	US-09-815-944-1
12	500	100.0	4768	13	US-09-811-361-1
13	500	100.0	4768	13	US-09-815-937-1
14	500	100.0	4768	13	US-09-885-816-1

```
15 500 100.0 4768 14 US-10-087-523-1 Sequence 1, Appli
16 500 100.0 5100 10 US-09-954-483A-14 Sequence 14, Appli
17 500 100.0 6148 10 US-09-954-483A-13 Sequence 13, Appli
18 497 99.4 6355 9 US-09-816-790-2 Sequence 2, Appli
19 497 99.4 6355 9 US-09-861-077-2 Sequence 2, Appli
20 497 99.4 6355 9 US-09-815-825-2 Sequence 2, Appli
21 497 99.4 6355 9 US-09-815-935-2 Sequence 2, Appli
22 497 99.4 6355 9 US-09-815-944-2 Sequence 2, Appli
23 497 99.4 6355 13 US-09-811-361-2 Sequence 2, Appli
24 497 99.4 6355 13 US-09-815-937-2 Sequence 2, Appli
25 497 99.4 6355 13 US-09-885-816-2 Sequence 2, Appli
26 497 99.4 6355 14 US-10-087-523-2 Sequence 2, Appli
27 496.8 99.4 3166 15 US-10-161-403-114 Sequence 114, App
28 496.8 99.4 4615 15 US-10-161-403-117 Sequence 117, App
29 496.8 99.4 6119 15 US-10-161-403-126 Sequence 126, App
30 496.8 99.4 7600 15 US-10-161-403-115 Sequence 115, App
31 496.8 99.4 7821 15 US-10-161-403-116 Sequence 116, App
32 496.8 99.4 8851 15 US-10-161-403-124 Sequence 124, App
33 496.8 99.4 9080 15 US-10-161-403-110 Sequence 110, App
34 496.8 99.4 10474 15 US-10-161-403-125 Sequence 125, App
35 492.4 98.5 8590 15 US-10-038-722-75 Sequence 75, Appli
36 492 98.4 6714 10 US-09-981-002-5 Sequence 5, Appli
37 491.6 98.3 2836 13 US-10-423-828-90 Sequence 90, Appli
38 491.4 98.3 1086 15 US-10-182-616-11 Sequence 11, Appli
39 491.4 98.3 2212 10 US-09-764-891-5577 Sequence 5577, Ap
40 491.4 98.3 2212 10 US-09-764-891-5607 Sequence 1456, Ap
41 491.4 98.3 2213 9 US-09-764-868-1456 Sequence 1462, Ap
42 491.4 98.3 2213 10 US-09-764-891-5565 Sequence 5565, Ap
43 491.4 98.3 2213 10 US-09-764-891-5570 Sequence 5570, Ap
44 491.4 98.3 2213 10 US-09-764-891-5578 Sequence 5578, Ap
45 491.4 98.3 2213 10 US-09-764-891-5578 Sequence 5578, Ap
```

ALIGNMENTS

RESULT 1

```
US-10-161-403-119
; Sequence 119, Application US/10161403
; Publication No. US20030119104A1
; GENERAL INFORMATION:
; APPLICANT: Perkins, Edward
; APPLICANT: Perez, Carl
; APPLICANT: Lindenbaum, Michael
; APPLICANT: Greene, Amy
; APPLICANT: Leung, Josephine
; APPLICANT: Fleming, Elena
; APPLICANT: Stewart, Sandra
; APPLICANT: Shellard, Joan
; TITLE OF INVENTION: CHROMOSOME-BASED PLATFORMS
; FILE REFERENCE: 24601-420
; CURRENT APPLICATION NUMBER: US/10161,403
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: 60/294,758
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: 60/366,891
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119
; LENGTH: 2814
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pLITMUS38 Plasmid
US-10-161-403-119
```

Query Match 100.0%; Score 500; DB 15; Length 2814;

Best Local Similarity 100.0%; Pred. No. 1.68-138;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTAACTAGTCAGTGGCAGCTTTTCGGGAAATGCGCGAACCCTATTGTGTTATT 60

Db 3568 ATACATTCAATATGATCCGCTCATGAGACATTAACCCCTGATTAATGCTTCAATATAT 3509
QY 128 TGAAGAGAGAGATGATGAGTATCAACATTTCCGTCGCGCTTATTCCTTTTGGG 187
Db 3508 TGAAGAGAGAGATGATGAGTATCAACATTTCCGTCGCGCTTATTCCTTTTGGG 3449
QY 188 GCATTTTGCCTTCCCTGTTTCTCACCAGAGAGCGTGTGAAAGTAAAGATGCTGAA 247
Db 3448 GCATTTTGCCTTCCCTGTTTCTCACCAGAGAGCGTGTGAAAGTAAAGATGCTGAA 3389
QY 248 GATCAGTTGGGTGACGAGTGGTGTACATCGAATCTCAACAGCGGTAAATCCTT 307
Db 3388 GATCAGTTGGGTGACGAGTGGTGTACATCGAATCTCAACAGCGGTAAATCCTT 3329
QY 308 GAGAGTTTTCGCGCGGAGAGCGTGTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 3328 GAGAGTTTTCGCGCGGAGAGCGTGTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 3269
QY 368 GCGCGGTATTAATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 427
Db 3268 GCGCGGTATTAATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 3209
QY 428 TCTCAGATGACTTGGTGGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 487
Db 3208 TCTCAGATGACTTGGTGGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 3149
QY 488 ACAGTAAGAGAAT 500
Db 3148 ACAGTAAGAGAAT 3136

RESULT 15

US-08-645-641-120/c
; Sequence 120, Application US/08645641
; Patent No. 5719032
; GENERAL INFORMATION:
; APPLICANT: Lenberg, Nile
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 5719032-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESS: William M. Smith
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,641
; FILING DATE: 20-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 14643-000913
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-645-641-120
Query Match 98.3%; Score 491.4; DB 1; Length 3699;
Best Local Similarity 99.8%; Pred. No. 6.2e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACCTCAGGTGGCACTTTTCGGGGAAATGCGCGGAAACCCCTATTGTTTATTTTCTAA 67
Db 3628 ACCTCAGGTGGCACTTTTCGGGGAAATGCGCGGAAACCCCTATTGTTTATTTTCTAA 3569
QY 68 ATACATTTCAATATGATATCCGCTCATGAGACAATTAACCTTGATTAATGTTCAATATAT 127
Db 3568 ATACATTTCAATATGATATCCGCTCATGAGACAATTAACCTTGATTAATGTTCAATATAT 3509
QY 128 TGAAGAGAGAGATGATGAGTATCAACATTTCCGTCGCGCTTATTCCTTTTGGG 187
Db 3508 TGAAGAGAGAGATGATGAGTATCAACATTTCCGTCGCGCTTATTCCTTTTGGG 3449
QY 188 GCATTTTGCCTTCCCTGTTTCTCACCAGAGAGCGTGTGAAAGTAAAGATGCTGAA 247
Db 3448 GCATTTTGCCTTCCCTGTTTCTCACCAGAGAGCGTGTGAAAGTAAAGATGCTGAA 3389
QY 248 GATCAGTTGGGTGACGAGTGGTGTACATCGAATCTCAACAGCGGTAAATCCTT 307
Db 3388 GATCAGTTGGGTGACGAGTGGTGTACATCGAATCTCAACAGCGGTAAATCCTT 3329
QY 308 GAGAGTTTTCGCGCGGAGAGCGTGTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
Db 3328 GAGAGTTTTCGCGCGGAGAGCGTGTCTCAATGATGAGCACTTTTAAAGTTCTGCTATGT 3269
QY 368 GCGCGGTATTAATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 427
Db 3268 GCGCGGTATTAATCCCGTGTGACCGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 3209
QY 428 TCTCAGATGACTTGGTGGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 487
Db 3208 TCTCAGATGACTTGGTGGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 3149
QY 488 ACAGTAAGAGAAT 500
Db 3148 ACAGTAAGAGAAT 3136

Search completed: April 22, 2004, 12:58:18
Job time : 90.5798 secs


```

Db      3147 ACAGTAAGAGAAT 3135
|||||
PCT-US92-06185-49/c
; Sequence 49, Application PC/TUS9206185
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic Non-Human Animals Capable of
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: William M. Smith
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06185
; FILING DATE: 19910828
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 87654
; REFERENCE/DOCKET NUMBER: 14643-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US92-06185-49

Query Match      98.3%; Score 491.4; DB 5; Length 3698;
Best Local Similarity 99.8%; Pred. No. 6.2e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      8  ACGTCAGGTGGCACTTTTCGGGGAAATGTCGCGGAACCCCTATTGTTTATTTCTAA 67
Db      3627 ACGTCAGGTGGCACTTTTCGGGGAAATGTCGCGGAACCCCTATTGTTTATTTCTAA 3568

QY      68  ATACATTCAAATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATATAT 127
Db      3567 ATACATTCAAATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATATAT 3508

QY      128 TGAAGAAGAGAGATGATGAGTATTCACATTCGCGTGCCTTATCCCTTTTTCG 187
Db      3507 TGAAGAAGAGAGATGATGAGTATTCACATTCGCGTGCCTTATCCCTTTTTCG 3448

QY      188 GCATTTTCCTTCCTGTTTTCCTGTCACCCGAAACGCTGTAAGTAAGATGCTGAA 247
Db      3447 GCATTTTCCTTCCTGTTTTCCTGTCACCCGAAACGCTGTAAGTAAGATGCTGAA 3388

QY      248 GATCAGTTGGTGACAGAGTGGGTATCATGAACTGATCTCAACACCGGTAAAGATCCTT 307
Db      3387 GATCAGTTGGTGACAGAGTGGGTATCATGAACTGATCTCAACACCGGTAAAGATCCTT 3328

QY      308 GAGAGTTTTCCTCCCGGAAGACGTTTCTCCATGATGAGCACTTTTAAAGTCTGCTATGT 367
Db      3327 GAGAGTTTTCCTCCCGGAAGACGTTTCTCCATGATGAGCACTTTTAAAGTCTGCTATGT 3268

```

```

QY      368  GGCGCGGTATTATCCCGTGTGACGCGGCAAGAGCACTCGGTGCGCGCATACACTAT 427
Db      3267 GGCGCGGTATTATCCCGTGTGACGCGGCAAGAGCACTCGGTGCGCGCATACACTAT 3208

QY      428  TCTCAGAAATGACITGGTTGAGTACTCACAGTCACAGAAAAGCATCTTACGGATGCGCATG 487
Db      3207 TCTCAGAAATGACITGGTTGAGTACTCACAGTCACAGAAAAGCATCTTACGGATGCGCATG 3148

QY      488  ACAGTAAGAGAAT 500
Db      3147 ACAGTAAGAGAAT 3135

RESULT 14
US-08-053-131-120/c
; Sequence 120, Application US/08053131
; Patent No. 5661016
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 200
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/053,131
; FILING DATE: 26-APR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 14643-9-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-053-131-120

Query Match      98.3%; Score 491.4; DB 1; Length 3699;
Best Local Similarity 99.8%; Pred. No. 6.2e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      8  ACGTCAGGTGGCACTTTTCGGGGAAATGTCGCGGAACCCCTATTGTTTATTTCTAA 67
Db      3628 ACGTCAGGTGGCACTTTTCGGGGAAATGTCGCGGAACCCCTATTGTTTATTTCTAA 3569

QY      68  ATACATTCAAATATGATCCGCTCATGAGACAATAACCCCTGATAAATGCTTCAATATAT 127

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/834,539A
FILING DATE: 1992-02-05
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 14643-5
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 3698 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-834-539A-49

Query Match 98.3%; Score 491.4; DB 1; Length 3698;
Best Local Similarity 99.8%; Pred No. 6.2e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACCTCAGTGGGACATTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 67
DB 3627 ACCTCAGTGGGACATTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 3568

QY 68 ATACATTCAAATATGATATCCGCTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 127
DB 3567 ATACATTCAAATATGATATCCGCTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 3508

QY 128 TGAAGAGGAGATGATGAGTATTCACATTTCCGTCGCGCCCTATTGTTTATTTTCTAA 187
DB 3507 TGAAGAGGAGATGATGAGTATTCACATTTCCGTCGCGCCCTATTGTTTATTTTCTAA 3448

QY 188 GCATTTTCGCTTCCTGTTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 67
DB 3447 GCATTTTCGCTTCCTGTTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 3568

QY 248 GATCAGTTGGGTGACAGTGGGTTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 127
DB 3567 GATCAGTTGGGTGACAGTGGGTTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 3508

QY 128 TGAAGAGGAGATGATGAGTATTCACATTTCCGTCGCGCCCTATTGTTTATTTTCTAA 187
DB 3507 TGAAGAGGAGATGATGAGTATTCACATTTCCGTCGCGCCCTATTGTTTATTTTCTAA 3448

QY 188 GCATTTTCGCTTCCTGTTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 67
DB 3447 GCATTTTCGCTTCCTGTTTTCGGGGAAATGTCGGCGAACCCTATTGTTTATTTTCTAA 3568

QY 248 GATCAGTTGGGTGACAGTGGGTTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 127
DB 3567 GATCAGTTGGGTGACAGTGGGTTCATGAGCAATACCCCTGATAATGCTTCAATAATAT 3508

QY 308 GAGAGTTTTCGCGCGAAGAACGTTCTCCATGATGAGCACTTTTAAAGTTCGCTATGT 367
DB 3327 GAGAGTTTTCGCGCGAAGAACGTTCTCCATGATGAGCACTTTTAAAGTTCGCTATGT 3268

QY 368 GCGCGGTATTATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 427
DB 3267 GCGCGGTATTATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGCGCATACACTAT 3208

QY 428 TCTCAGATGACTTGGTTGAGTACTACAGTCAAGAAAGCATCTTACGGATGGCATG 487
DB 3207 TCTCAGATGACTTGGTTGAGTACTACAGTCAAGAAAGCATCTTACGGATGGCATG 3148

QY 488 ACAGTAAGAGAAAT 500
DB 3147 ACAGTAAGAGAAAT 3135

RESULT 12
US-08-800-353-49/c
Sequence 49, Application US/08800353
Patent No. 5874299
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 5874299-Human Animals Capable of Producing Heterologous Antibodies

RESULT 10
US-09-058-483-9/c
; Sequence 9, Application US/09058483A
; Patent No. 6365347
; GENERAL INFORMATION:
; APPLICANT: Murray, Andrew W.
; APPLICANT: Smith, Dana L.
; APPLICANT: Sorger, Peter K.
; APPLICANT: No. 6365347man, Thea C.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING DISRUPTORS OF BIOLOGICAL

RESULT 11
US-07-834-539A-49/c
Sequence 49, Application US/07834539A
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
APPLICANT: Kay, Robert M.
TITLE OF INVENTION: Transgenic Non-Human Animals Capable of
Producing Heterologous Antibodies
TITLE OF INVENTION: 77
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: William M. Smith
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

QY 308 GAGAGTTTTCGCCCGAAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
DB 1538 GAGAGTTTTCGCCCGAAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGT 1479
QY 368 GCGCGGTTATTATCCCGTGTGACCGCGGGAAGAGCACTCGGTGCGCGGATACACTAT 427
DB 1478 GCGCGGTTATTATCCCGTGTGACCGCGGGAAGAGCACTCGGTGCGCGGATACACTAT 1419
QY 428 TCTCAGATGACTTGGTTGAGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 487
DB 1418 TCTCAGATGACTTGGTTGAGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 1359
QY 488 ACAGTAAGAGAAAT 500
DB 1358 ACAGTAAGAGAAAT 1346

RESULT 5

US-08-906-957-9/C
; Sequence 9, Application US/08906957
; Patent No. 5856142
; GENERAL INFORMATION:
; APPLICANT: LEGOUX, Richard
; APPLICANT: MALDONADO, Paul
; APPLICANT: SALOME, Marc
; TITLE OF INVENTION: Method for the extraction of
; TITLE OF INVENTION: periplasmic proteins of prokaryotic microorganisms in the
; TITLE OF INVENTION: presence of arginine
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bacon & Thomas
; STREET: 625 Slaters Lane - Fourth Floor
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,957
; FILING DATE: 06-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/594,469
; FILING DATE:
; APPLICATION NUMBER: FR 95 01083
; FILING DATE: 31-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FICHTER, Richard E
; REFERENCE/DOCKET NUMBER: 26,382
; TELEPHONE: (703) 683-0500
; TELEFAX: (703) 683-1080
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1905 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-906-957-9

Query Match 98.3%; Score 491.4; DB 2; Length 1905;
Best Local Similarity 99.8%; Pred. No. 4.6e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 8 ACCTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 67
DB 1838 ACCTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 1779

QY 68 ATACATTCAAATATGTATCCGTCATGAGACAATACCCCTGATAAATGCTTCAATAATAT 127
DB 1778 ATACATTCAAATATGTATCCGTCATGAGACAATACCCCTGATAAATGCTTCAATAATAT 1719
QY 128 TGAATAAGGAAGAGATGATGATATTCACATTTCCGTTGCGCCCTTATTTCCCTTTTTCGG 187
DB 1718 TGAATAAGGAAGAGATGATGATATTCACATTTCCGTTGCGCCCTTATTTCCCTTTTTCGG 1659
QY 188 GCATTTTGGCTTTCCTGTTTTTTCCTCACCAGAAAACGCTGGTGAAGTAAAAGATCCTGAA 247
DB 1658 GCATTTTGGCTTTCCTGTTTTTTCCTCACCAGAAAACGCTGGTGAAGTAAAAGATCCTGAA 1599
QY 248 GATCAGTTGGGTGCGAGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 307
DB 1598 GATCAGTTGGGTGCGAGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTT 1539
QY 308 GAGAGTTTTTCGCCCGAAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGT 367
DB 1538 GAGAGTTTTTCGCCCGAAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGT 1479
QY 368 GCGCGGTTATTATCCCGTGTGACCGCGGGAAGAGCACTCGGTGCGCGGATACACTAT 427
DB 1478 GCGCGGTTATTATCCCGTGTGACCGCGGGAAGAGCACTCGGTGCGCGGATACACTAT 1419
QY 428 TCTCAGATGACTTGGTTGAGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 487
DB 1418 TCTCAGATGACTTGGTTGAGTACTCACCAGTTCACAGAAAGCATCTTACGGATGGCATG 1359
QY 488 ACAGTAAGAGAAAT 500
DB 1358 ACAGTAAGAGAAAT 1346

RESULT 6

US-09-202-904A-13
; Sequence 13, Application US/09202904A
; Patent No. 6395471
; GENERAL INFORMATION:
; APPLICANT: Kang, Sung Key
; APPLICANT: Hahn, Bumsuk
; TITLE OF INVENTION: Hepatitis C Surrogate Virus for Testing
; TITLE OF INVENTION: the Activity of Hepatitis C Virus Protease, a Recombinant
; FILE OF INVENTION: Gene and a Use Thereof
; FILE REFERENCE: A32210-PCT-USA 072944.0104
; CURRENT APPLICATION NUMBER: US/09/202,904A
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/KR97/00120
; PRIOR FILING DATE: 1997-06-25
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Genetically engineered virus derived from
; OTHER INFORMATION: poliovirus and hepatitis C virus
US-09-202-904A-13

Query Match 98.3%; Score 491.4; DB 4; Length 2320;
Best Local Similarity 99.8%; Pred. No. 5e-137;
Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 ACCTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 67
DB 1256 ACCTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATTCTTAA 1315
QY 68 ATACATTCAAATATGTATCCGTCATGAGACAATACCCCTGATAAATGCTTCAATAATAT 127
DB 1316 ATACATTCAAATATGTATCCGTCATGAGACAATACCCCTGATAAATGCTTCAATAATAT 1375
QY 128 TGAATAAGGAAGAGATGATGATATTCACATTTCCGTTGCGCCCTTATTTCCCTTTTTCGG 187

```

; MEDIUM TYPE: Floppy disk, 5.25", 360 kb.
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in 1.0, V1.25; EDIX; Wordperfect.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,623C
; FILING DATE: 19-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Blosser, G. Harley
; REGISTRATION NUMBER: 33,650
; REFERENCE/DOCKET NUMBER: WNB4900
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314/231-5400
; TELEFAX: 314/231-4342
; TELEX: 6502697583 MCI
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Expression vector
; IMMEDIATE SOURCE:
; CLONE: pWB254b
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1665
US-08-021-623C-5

```

```

Query Match      98.4%; Score 492; DB 1; Length 6714;
Best Local Similarity 99.0%; Pred. No. 5.3e-137; Mismatches 5; Indels 0; Gaps 0;
Matches 495; Conservative 0;

QY 1 GTTAACACGTCAGTGGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 60
DB 2269 GTTACAATTTTCAGTGGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 2328
QY 61 TTTCCTAATACATTCARATATGATCCGCTCATGAGACAATACCCCTGATAAATGCTTCA 120
DB 2329 TTTCCTAATACATTCARATATGATCCGCTCATGAGACAATACCCCTGATAAATGCTTCA 2388
QY 121 ATAATATTGAAAGGAAGAGATGATGATTCAACATTTCCGCTGTCGCCCTTATTCCTT 180
DB 2389 ATAATATTGAAAGGAAGAGATGATGATTCAACATTTCCGCTGTCGCCCTTATTCCTT 2448
QY 181 TTTTCGGGCAATTTGCGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
DB 2449 TTTTCGGGCAATTTGCGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 2508
QY 241 TGCTGAAGATCAGTTGGGTGACAGTGGGTATCATCGAACTGGATCTCAACAGCGGTAA 300
DB 2509 TGCTGAAGATCAGTTGGGTGACAGTGGGTATCATCGAACTGGATCTCAACAGCGGTAA 2568
QY 301 GATCCTTGAGAGTTTTCGCCCGGAAGACGTTCTCCAATGATGAGCATTCTTAAGTCT 360
DB 2569 GATCCTTGAGAGTTTTCGCCCGGAAGACGTTTTCCTCAATGATGAGCATTCTTAAGTCT 2628
QY 361 GCTATGTGGCGGCTATTATCCGCTGTTGAGCGCGGGAAGAGCACTCGGTCGCCGAT 420
DB 2629 GCTATGTGGCGGCTATTATCCGCTGTTGAGCGCGGGAAGAGCACTCGGTCGCCGAT 2688
QY 421 ACATATTCTCAGAAATGATCTGGTTGAGTATCTCACCAGTCTCAGAAAGCATCTTACGGA 480
DB 2689 ACATATTCTCAGAAATGATCTGGTTGAGTATCTCACCAGTCTCAGAAAGCATCTTACGGA 2748
QY 481 TGGCATGACGTAAGAGAT 500
DB 2749 TGGCATGACGTAAGAGAT 2768

```

RESULT 4

```

US-08-594-469-9/c
; Sequence 9, Application US/08594469
; Patent No. 5700665
; GENERAL INFORMATION:
; APPLICANT: LEGOUX, Richard
; APPLICANT: MALDONADO, Paul
; APPLICANT: SALOME, Marc
; TITLE OF INVENTION: Method for the extraction of
; TITLE OF INVENTION: periplasmic proteins of prokaryotic microorganisms in the
; TITLE OF INVENTION: presence of arginine
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bacon & Thomas
; STREET: 625 Slaters Lane - Fourth Floor
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/594,469
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 95 01083
; FILING DATE: 31-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FICHTER, Richard E
; REGISTRATION NUMBER: 26,382
; REFERENCE/DOCKET NUMBER: REF/LEGOUX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 683-0500
; TELEFAX: (703) 683-1080
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1905 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-594-469-9

```

```

Query Match      98.3%; Score 491.4; DB 1; Length 1905;
Best Local Similarity 99.8%; Pred. No. 4.6e-137; Mismatches 1; Indels 0; Gaps 0;
Matches 492; Conservative 0;

QY 8 ACCTCAGGTGGCAGCTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTGTTTCTAA 67
DB 1838 ACCTCAGGTGGCAGCTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATTGTTTCTAA 1779
QY 68 ATACATTCAAATATGATCCGCTCATGAGACAATACCCCTGATTAATGCTTCAATATAT 127
DB 1778 ATACATTCAAATATGATCCGCTCATGAGACAATACCCCTGATTAATGCTTCAATATAT 1719
QY 128 TGAAGAAGGAAGAGATGATGATTCAACATTTCCGCTGTCGCCCTTATTCCTTTTTCGG 187
DB 1718 TGAAGAAGGAAGAGATGATGATTCAACATTTCCGCTGTCGCCCTTATTCCTTTTTCGG 1659
QY 188 GCATTTTGGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGTCTGAA 247
DB 1658 GCATTTTGGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGTCTGAA 1599
QY 248 GATCAGTGGGTGACAGAGTGGGTACATCGAACTGGATCTCAACAGCGGTGAAGTCTT 307
DB 1598 GATCAGTGGGTGACAGAGTGGGTACATCGAACTGGATCTCAACAGCGGTGAAGTCTT 1539

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:26:21 ; Search time 88.5798 Seconds
(without alignments)
3132.491 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500

Perfect score: 500

Sequence: 1 gttactacgtcaggtgga.....tgccatgacagtaagaat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	500	100.0	3476	US-09-380-484A-1	Sequence 1, Appli
C 2	491.4	98.5	8590	US-08-358-160-70	Sequence 70, Appl
C 3	492.4	98.4	6714	US-08-021-623C-5	Sequence 5, Appli
C 4	491.4	98.3	1905	US-08-594-469-9	Sequence 9, Appli
C 5	491.4	98.3	1905	US-08-906-957-9	Sequence 9, Appli
C 6	491.4	98.3	2320	US-09-202-904A-13	Sequence 13, Appl
C 7	491.4	98.3	3122	US-09-042-353-152	Sequence 152, App
C 8	491.4	98.3	3122	US-08-758-417A-416	Sequence 416, App
C 9	491.4	98.3	3418	US-08-944-916-12	Sequence 12, Appl
C 10	491.4	98.3	3516	US-09-058-483-9	Sequence 9, Appli
C 11	491.4	98.3	3698	US-07-834-539A-49	Sequence 49, Appl
C 12	491.4	98.3	3698	US-08-800-353-49	Sequence 49, Appl
C 13	491.4	98.3	3698	PCT-US92-06185-49	Sequence 49, Appl
C 14	491.4	98.3	3699	US-08-053-131-120	Sequence 120, App
C 15	491.4	98.3	3699	US-08-645-641-120	Sequence 120, App
C 16	491.4	98.3	3699	US-07-853-408B-120	Sequence 120, App
C 17	491.4	98.3	3699	US-08-096-762-120	Sequence 120, App
C 18	491.4	98.3	3699	US-08-308-865-120	Sequence 120, App
C 19	491.4	98.3	3699	PCT-US92-10983-120	Sequence 120, App
C 20	491.4	98.3	3754	US-08-586-740A-6	Sequence 6, Appli
C 21	491.4	98.3	3754	US-08-379-611-17	Sequence 17, Appl
C 22	491.4	98.3	3769	US-08-379-611-18	Sequence 18, Appl
C 23	491.4	98.3	4366	US-08-586-740A-12	Sequence 12, Appl
C 24	491.4	98.3	4378	US-08-586-740A-9	Sequence 9, Appli
C 25	491.4	98.3	4410	US-08-594-469-1	Sequence 1, Appli
C 26	491.4	98.3	4410	US-08-906-957-1	Sequence 1, Appli
C 27	491.4	98.3	4451	US-09-303-064-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1

```
US-09-380-484A-1/c
; Sequence 1, Application US/09380484A
; Patent No. 6632928
; GENERAL INFORMATION:
; APPLICANT: Neville, David M.
; APPLICANT: Knechtie, Stuart
; APPLICANT: Thomas, Judith M.
; APPLICANT: Thompson, Jerry T.
; APPLICANT: Hu, Huaizhong
; APPLICANT: Ma, Shenglin
; TITLE OF INVENTION: IMMUNOTOXINS AND METHODS OF INDUCING
; FILE REFERENCE: 14028.0287
; CURRENT APPLICATION NUMBER: US/09/380,484A
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: PCT/US98/04303
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: 60/039,987
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3476
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note =
; OTHER INFORMATION: synthetic construct
; OTHER INFORMATION: Description of Artificial Sequence:/note =
; OTHER INFORMATION: synthetic
US-09-380-484A-1

Query Match 100.0%; Score 500; DB 4; Length 3476;
Best Local Similarity 100.0%; Pred. No. 1.6e-139;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACACTGCTCAGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 60
3476 GTTAACACTGCTCAGTGGCACTTTTCGGGAAATGTCGGGAAACCCCTATTGTTTATT 3417

QY 61 TTCTTAATATACATTCAAAATATGTCCTCATGAGACAAATAACCTGTAATGCTTCA 120
3416 TTCTTAATATACATTCAAAATATGTCCTCATGAGACAAATAACCTGTAATGCTTCA 3357

QY 121 ATAAATTTGAAAAGAGAGATGAGTATTCACATTTCGGTGTGCGCCCTATTCCCTT 180
3356 ATAAATTTGAAAAGAGAGATGAGTATTCACATTTCGGTGTGCGCCCTATTCCCTT 3297

QY 181 TTTTGGGCACTTTTCGCTTCCTGTTTTTTCCTACCCAGAACGCTGGTGAAGTAAAGA 240
```

CC transgenic animals containing targeted gene modifications. This sequence
 CC represents the positive selection vector construct c3408
 XX
 SQ Sequence 5759 BP; 1274 A; 1526 C; 1658 G; 1301 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 6; Length 5759;
 Best Local Similarity 100.0%; Pred. NO. 4.4e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACTACGTACAGTGGACATTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 60
 Db 3296 GTTAACTACGTACAGTGGACATTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 3355

QY 61 TTCTAAATACATTCATAATATGTATCCGCTCATGAGACAAATACCCCTGATAAATGCTTCA 120
 Db 3356 TTCTAAATACATTCATAATATGTATCCGCTCATGAGACAAATACCCCTGATAAATGCTTCA 3415

QY 121 ATAAATATGAAAAGAGAGATGATTCACATTTCCGTGTCGCCCTTATTCCTTT 180
 Db 3416 ATAAATATGAAAAGAGAGATGATTCACATTTCCGTGTCGCCCTTATTCCTTT 3475

QY 181 TTTTGGCGATTTTGCTTCCTTCTGTTTTTGTCTCACCAGAAACGCTGGTGAAGTAAAGA 240
 Db 3476 TTTTGGCGATTTTGCTTCCTTCTGTTTTTGTCTCACCAGAAACGCTGGTGAAGTAAAGA 3535

QY 241 TGCTGAAGATCAGTTGGTGCACGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAA 300
 Db 3536 TGCTGAAGATCAGTTGGTGCACGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAA 3595

QY 301 GATCCTTGAGAGTTTTTCGCCCGGAGACAGACGTTCTCCAATGATGAGACATTTTAAAGTTCT 360
 Db 3596 GATCCTTGAGAGTTTTTCGCCCGGAGACAGACGTTCTCCAATGATGAGACATTTTAAAGTTCT 3655

QY 361 GCTATGTCGGCGGTATTATCCGTTGTCACCGCGGCAAGAGCAACTCGGTCCCGCAT 420
 Db 3656 GCTATGTCGGCGGTATTATCCGTTGTCACCGCGGCAAGAGCAACTCGGTCCCGCAT 3715

QY 421 ACACTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTTACGGA 480
 Db 3716 ACACTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTTACGGA 3775

QY 481 TGGCATGACAGTAAGAGAT 500
 Db 3776 TGGCATGACAGTAAGAGAT 3795

Search completed: April 22, 2004, 10:42:35
 Job time : 436.846 secs

```

RESULT 14
AAD28659
ID AAD28659 standard; DNA; 4768 BP.
XX
AC AAD28659;
XX
DT 07-MAY-2002 (first entry)
XX
DE Plasmid pDG2 vector.
XX
KW Plasmid pDG2; DNA construct; embryonic stem cell; cell disruption; Neo';
XX neomycin; ampicillin resistance gene; ds.
XX
OS Unidentified.
XX
PN WO200204621-A2.
XX
PD 17-JAN-2002.
XX
PF 11-JUL-2000; 2000WO-US018812.
XX
PR 11-JUL-2000; 2000WO-US018812.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Klein RD, Brennan TU;
XX
DR WPI; 2002-164642/21.
XX
DE Novel nucleotide construct for generating DNA constructs for introducing
PT into embryonic stem cell, comprising a sequence encoding a positive
PT selection marker flanked by restriction enzyme sites.
XX
PS Claim 10; Fig 2B; 64pp; English.
XX
CC The invention relates to nucleotide construct for generating DNA
CC constructs. The nucleotide construct comprises a sequence encoding a
CC positive selection marker flanked by restriction enzyme sites, where
CC restriction site is flanked by sequences which are not complementary to
CC each other and which do not include at least one type of base at any
CC position, where the construct can be treated so that single-stranded
CC regions are created at each sequence lacking at least one nucleotide. The
CC nucleotide construct is useful in a rapid and efficient method for
CC generating DNA constructs suitable for introduction into embryonic stem
CC cells and for disrupting the function of a gene in a cell. The present
CC sequence is plasmid pDG2 vector construct containing an ampicillin
CC resistance gene and neomycin gene (Neo'). On each site of the Neo' gene
CC are two sites for ligation independent cloning along with restriction
CC sites
XX
SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;
Query Match 100.0%; Score 500; DB 6; Length 4768;
Best Local Similarity 100.0%; Fred. No. 4.2e-124;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACACGTCAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 60
DB 1 GTTAACACGTCAGTGGCACTTTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 60
QY 61 TTCTTAATACATTCAAATATGATTCGCTCATGAGACATAACCCGTGTAATGCTTCA 120
DB 61 TTCTTAATACATTCAAATATGATTCGCTCATGAGACATAACCCGTGTAATGCTTCA 120
QY 121 ATAATATTGAAAAAGGAGATGATGATTCACATTTTCGTCGCCCTTATTCGCTT 180
DB 121 ATAATATTGAAAAAGGAGATGATGATTCACATTTTCGTCGCCCTTATTCGCTT 180
QY 191 TTTTGGGCATTTTGCCTTCTGTTTTCGTCACCCAGAACCGTGTGGAAGTAAAGA 240
DB 191 TTTTGGGCATTTTGCCTTCTGTTTTCGTCACCCAGAACCGTGTGGAAGTAAAGA 240
QY 241 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAATCGATCTCACACGGGTAA 300

```

```

Db 241 TGCTGAAGATCAGTTGGGTGCACGAGTGGTTCACATCGAATCGGATCTCAACAGCGGTAA 300
QY 301 GATCCTTTGAGAGTTTTCGCCCGGAGAACGTTCTTCCAATGATGAGCACATTTTAAAGTTCT 360
Db 301 GATCCTTTGAGAGTTTTCGCCCGGAGAACGTTCTTCCAATGATGAGCACATTTTAAAGTTCT 360
QY 361 GCTATGTCGGCGGTATTATCCCGTGTTCACGCCGGGCAAGAGCAACTCGGTCCGCGCAT 420
Db 361 GCTATGTCGGCGGTATTATCCCGTGTTCACGCCGGGCAAGAGCAACTCGGTCCGCGCAT 420
QY 421 ACACTATTCTCAGAATGACTTGTGTTGAGTACTCACCAGTCACAGAAAACATCTTACCGA 480
Db 421 ACACTATTCTCAGAATGACTTGTGTTGAGTACTCACCAGTCACAGAAAACATCTTACCGA 480
QY 481 TGGCATGACAGTAAGAGAAAT 500
Db 481 TGGCATGACAGTAAGAGAAAT 500

RESULT 15
ABK49521
ID ABK49521 standard; DNA; 5759 BP.
XX
AC ABK49521;
XX
DT 15-JUL-2002 (first entry)
XX
DE Positive selection vector construct c3408.
XX
KW Transgenic animal; targeting vector; positive selection vector;
KW homologous recombination; target gene modification; transgenic animal;
KW c3408; ds.
XX
OS Synthetic.
XX
PN WO200222834-A2.
XX
PD 21-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-US028892.
XX
PR 15-SEP-2000; 2000US-0232957P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Siebel C, Brennan TU;
XX
DR WPI; 2002-383132/41.
XX
PT Novel targeting vector modifying target gene, has first and second
PT sequences homologous to target gene portions, a selectable marker
PT cassette and regulator, useful for producing animals with targeted gene
PT modifications.
XX
PS Example 1; Fig 14A-B; 43pp; English.
XX
CC The invention describes a targeting vector (positive selection vector)
CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
CC and S2) homologous to a portion or region of a target gene, a selectable
CC marker cassette and a regulator. (I) is useful for producing cells
CC comprising a modification of the target gene which involves introducing
CC (I) into cells capable of homologous recombination, selecting for cells
CC expressing the selectable marker and identifying cells containing the
CC modification of the target gene. Use of (I) for enriching cells
CC comprising disruption or modification of target gene enhances recovery of
CC cells having targeting vector integrated via homologous recombination
CC into the genomes of the cells. (I) is capable of modifying target gene in
CC a cell with high efficiency and specificity. Use of (I) provides a faster
CC and more efficient means for isolating and selecting cells comprising
CC target gene modification. Also use of (I) provides an increase over
CC previous technologies in both the speed and frequency at which homologous
CC recombination events can be recovered. (I) is also useful for creation of

```

Db 181 TTTTGGGCAATTTTGGCTTCTCTGTTTGTCTACCCAGAAACGCTGGTGAAGTAAAGA 240
Qy 241 TGGTGAAGATCAGTTGGTGGCAGAGTGGGTTTACATCGAACTGGATCTCAACAGCGGTAA 300
Db 241 TGGTGAAGATCAGTTGGTGGCAGAGTGGGTTTACATCGAACTGGATCTCAACAGCGGTAA 300
Qy 301 GATCCTTTGAGAGTTTTCGCCCCGAGAAACGTTCTCCAAATGATGAGCAGCTTTTAAAGTTCT 360
Db 301 GATCCTTTGAGAGTTTTCGCCCCGAGAAACGTTCTCCAAATGATGAGCAGCTTTTAAAGTTCT 360
Qy 361 GCTATGTTGGCGGCTATATCCCGTGTGTAGCGCGGCGGCAAGCACTGGTGGCGCAT 420
Db 361 GCTATGTTGGCGGCTATATCCCGTGTGTAGCGCGGCGGCAAGCACTGGTGGCGCAT 420
Qy 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAGCATCTTACGGA 480
Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAGCATCTTACGGA 480
Qy 481 TGGCATGACAGTAAGAGAT 500
Db 481 TGGCATGACAGTAAGAGAT 500

RESULT 13
AAS17143
ID AAS17143 standard; DNA; 4768 BP.
XX AC AAS17143;
XX
XX 14-FEB-2002 (first entry)
DT Gene targeting vector pDG2.
DE
XX
XX pDG2; ds; retina-specific nuclear receptor; gene targeting;
KW lymphoid-specific GPCR; melanocyte stimulating hormone receptor;
KW magnesium-dependent protein phosphatase; transgenic animal;
KW chemokine receptor 1-like protein; cGMP phosphodiesterase;
KW sulfolipase gene; tumour; cancer; retinal degeneration;
KW retinitis pigmentosa.
XX
XX Escherichia coli.
OS Synthetic.
XX
XX WO200167855-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US008664.
XX
XX 16-MAR-2000; 2000US-0190348P.
XX
XX 22-MAR-2000; 2000US-0191128P.
XX
XX 22-MAR-2000; 2000US-0191129P.
XX
XX 22-MAR-2000; 2000US-0191142P.
XX
XX 22-MAR-2000; 2000US-0191235P.
XX
XX 22-MAR-2000; 2000US-0191236P.
XX
XX 22-MAR-2000; 2000US-0191240P.
XX
XX 15-MAY-2000; 2000US-0204227P.
XX
XX 15-MAY-2000; 2000US-0204230P.
XX
XX 29-JUN-2000; 2000US-0215214P.
XX
XX 06-JUL-2000; 2000US-0216249P.
XX
XX 06-JUL-2000; 2000US-0216264P.
XX
XX 06-JUL-2000; 2000US-0216765P.
XX
XX 12-JUL-2000; 2000US-0218075P.
XX
XX 19-JUL-2000; 2000US-0219167P.
XX
XX 19-JUL-2000; 2000US-0219182P.
XX
XX 27-JUL-2000; 2000US-0221485P.
XX
XX 07-AUG-2000; 2000US-0223173P.
XX
XX (DELT-) DELTAGEN INC.
XX
XX Allen KD, Guenther C, Phillips R;
PI
XX

WPI; 2002-041167/05.
XX
XX New targeting construct comprising a first and a second polynucleotide
PT homologous to a target gene, and a selectable marker, useful for
PT introducing targeted mutations into embryonic cells.
XX
XX Example 3; Fig 2B; 105pp; English.
XX
XX The invention relates to a targeting construct comprising two sequences
CC homologous to a target gene, and a selectable marker, is new. The target
CC gene is a retina-specific nuclear receptor gene, a lymphoid-specific GPCR
CC (G protein coupled receptor) gene, a melanocyte stimulating hormone
CC receptor gene, a magnesium-dependent protein phosphatase gene, chemokine
CC receptor 1-like protein gene, a cGMP phosphodiesterase gene, or a
CC sulfolipase gene. Also included are transgenic mice comprising a
CC disruption in a target gene, where the mouse exhibits an eye abnormality,
CC cellular infiltration, hypochromic behaviour, lung abnormality, elevated
CC white blood cell count, abnormality in the aorta, kidney, liver, lymph
CC nodes, skin or salivary gland, increased body and organ weight, or
CC elevated levels of ALT (not defined), phosphorus, potassium, or
CC bilirubin, aggressive, hyperactive, increased activity or decreased
CC anxiety behaviour. The construct is used for introducing targeted
CC mutations into embryonic cells. The animal and cell-based systems may be
CC used as models for diseases or conditions associated with physiological,
CC histological or behavioural phenotypes relating to a disruption in a
CC target gene (e.g. tumours, cancer, retinal degeneration and retinitis
CC pigmentosa) and in screening or identifying compounds capable of
CC ameliorating or treating diseases. The present sequence is the vector
XX pDG2 used to generate the gene targeting construct of the invention
XX
XX Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 500; DB 6; Length 4768;
Best Local Similarity 100.0%; Pred. No. 4.2e-124;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTAACACTAGTCAGTGGCAGCTTTTCGGGGAATGTGCGGGAACCCCTATTTCCTTATT 60
Db 1 GTTAACACTAGTCAGTGGCAGCTTTTCGGGGAATGTGCGGGAACCCCTATTTCCTTATT 60
Qy 61 TTTCTAAATACATTCAAATATGTATCCGCTCATGAGCAATAACCCCTGATAAATGCTTCA 120
Db 61 TTTCTAAATACATTCAAATATGTATCCGCTCATGAGCAATAACCCCTGATAAATGCTTCA 120
Qy 121 ATAATATTGAAAAAGGAAGAGATGAGTATTCAACATTTCCGTTGCGCCCTTATTCCTT 180
Db 121 ATAATATTGAAAAAGGAAGAGATGAGTATTCAACATTTCCGTTGCGCCCTTATTCCTT 180
Qy 181 TTTTGGCGCATTTTGCCTTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Db 181 TTTTGGCGCATTTTGCCTTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
Qy 241 TGGTGAAGATCAGTTGGGTTGACAGAGTGGGTTTACATCGAACTGGATCTCAACAGCGGTAA 300
Db 241 TGGTGAAGATCAGTTGGGTTGACAGAGTGGGTTTACATCGAACTGGATCTCAACAGCGGTAA 300
Qy 301 GATCCTTTGAGAGTTTTCGCCCCGAGAAACGTTCTCCAAATGATGAGCAGCTTTTAAAGTTCT 360
Db 301 GATCCTTTGAGAGTTTTCGCCCCGAGAAACGTTCTCCAAATGATGAGCAGCTTTTAAAGTTCT 360
Qy 361 GCTATGTTGGCGGCTATATCCCGTGTGTAGCGCGGCGGCAAGCACTGGTGGCGCAT 420
Db 361 GCTATGTTGGCGGCTATATCCCGTGTGTAGCGCGGCGGCAAGCACTGGTGGCGCAT 420
Qy 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAGCATCTTACGGA 480
Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAGCATCTTACGGA 480
Qy 481 TGGCATGACAGTAAGAGAT 500
Db 481 TGGCATGACAGTAAGAGAT 500

CC The present sequence represents vector pDG2. This vector contains an
 CC ampicillin resistance gene and a neomycin gene. The vector is used in the
 CC invention. The specification describes a non-human transgenic animal
 CC comprising a disruption in the matrix metalloproteinase (MMP)-23 gene.
 CC Transgenic animals of the invention comprising a homozygous or
 CC heterozygous disruption in MMP23 gene are useful for identifying agents
 CC which modulate MMP23 expression or function. They are also useful for
 CC identifying agents that are capable of ameliorating a phenotype of a
 CC transgenic animal comprising a disruption in an MMP-23 gene or
 CC ameliorating a disease associated with the phenotype of a transgenic
 CC animal comprising a disruption in the MMP-23 gene. The animals are useful
 CC as an animal model for diseases, disorders and conditions characterized
 CC by a disruption in a gene encoding a metalloproteinase, more particularly
 CC disease, disorders and conditions associated with the phenotypes
 CC demonstrated by the knockout mice. The transgenic animals are useful as
 CC test substrates for identification of drugs, pharmaceuticals and
 CC therapies effective in treating diseases, disorders and conditions
 CC associated with disruption in the target gene. The animal is useful for
 CC testing and developing new treatments relating to behavioural phenotypes
 CC demonstrated by the animal models
 XX
 SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 6; Length 4768;
 Best Local Similarity 100.0%; Pred. No. 4.2e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTTAACACGTCAGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
 DB 1 GTTAACACGTCAGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
 QY 61 TTCTTAATACATTCAAATATGTATCGCTCATGAGACATAACCTGATAATGCTTCA 120
 DB 61 TTCTTAATACATTCAAATATGTATCGCTCATGAGACATAACCTGATAATGCTTCA 120
 QY 121 ATAAATTTGAAAAGGAAGAGTATGATGATTCACATTTCCGTCGCGCTTATTCCTT 180
 DB 121 ATAAATTTGAAAAGGAAGAGTATGATGATTCACATTTCCGTCGCGCTTATTCCTT 180
 QY 181 TTTTGGCGCATTTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 240
 DB 181 TTTTGGCGCATTTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 240
 QY 241 TGCTGAACATCAGTTGGGTGACGAGTGGGTATCATGCACTGATCTCAACGCGTAA 300
 DB 241 TGCTGAACATCAGTTGGGTGACGAGTGGGTATCATGCACTGATCTCAACGCGTAA 300
 QY 301 GATCCTTCAGAGTTTTCGCCCGGAAGACGTTCTCCAATGATGACACTTTTAAAGTCT 360
 DB 301 GATCCTTCAGAGTTTTCGCCCGGAAGACGTTCTCCAATGATGACACTTTTAAAGTCT 360
 QY 361 GCTATGTCGCGGTATTATCCCGTGTGACGCGGCAAGACGACTCGGTGCGCGCAT 420
 DB 361 GCTATGTCGCGGTATTATCCCGTGTGACGCGGCAAGACGACTCGGTGCGCGCAT 420
 QY 421 ACACATTTCTCAGATGACTTGGTTGAGTACTTCCAGTCTCAGACAGAAAGCATCTTACGGA 480
 DB 421 ACACATTTCTCAGATGACTTGGTTGAGTACTTCCAGTCTCAGACAGAAAGCATCTTACGGA 480
 QY 481 TGGCATGACAGTAAGAAAT 500
 DB 481 TGGCATGACAGTAAGAAAT 500

RESULT 12
 ABS53351
 ID ABS53351 standard; DNA; 4768 BP.
 XX
 AC ABS53351;
 XX
 XX 15-NOV-2002 (first entry)
 XX
 DE Plasmid vector pDG2 DNA sequence.

XX Nucleotide construct; positive selection marker; restriction enzyme site;
 KW embryonic stem cell; knock-out vector; genomic clone; mapping; PDG2;
 KW plasmid vector; ds.
 XX
 OS Synthetic.
 XX
 PN US2002086369-A1.
 XX
 PD 04-JUL-2002.
 XX
 XX 19-JUN-2001; 2001US-00895816.
 XX
 PR 17-NOV-1997; 97US-0084194P.
 PR 11-MAY-1998; 98US-0084949P.
 PR 17-NOV-1998; 98US-00133834.
 XX
 XX (DELT-) DELTAGEN INC.
 XX
 XX Klein RD, Brennan TJ;
 XX
 XX WPI; 2002-635678/68.
 DR
 XX
 PT Novel nucleotide construct useful for disrupting function of gene in
 PT embryonic stem cell, comprises sequence encoding positive selection
 PT marker flanked by restriction enzyme sites, to create a single-stranded
 PT region.
 XX
 PS Claim 10; Fig 2A; 37pp; English.
 XX
 CC The present invention relates to a new nucleotide construct comprising a
 CC sequence encoding positive selection marker flanked by restriction enzyme
 CC sites. The restriction enzyme sites are flanked by sequences which are
 CC not complementary to each other and which do not include at least one
 CC type of base at any position, where the construct is treated so that
 CC single-stranded regions are created at each sequence lacking at least one
 CC nucleotide. The invention is useful for disrupting the function of a
 CC target sequence or gene in a cell e.g. embryonic stem cell, by inserting
 CC sequences homologous to the target gene into the construct such that the
 CC sequences homologous to the target gene flank the positive selection
 CC marker, to produce a targeting construct, and introducing the targeting
 CC construct into the cell to produce a homologous recombinant and thus the
 CC function of the target gene or sequence is disrupted. The homologous
 CC sequences are sequences flanking the site in the target gene that is to
 CC be disrupted. The invention eliminates the need for the traditional
 CC hybridisation isolation of a single genomic clone, restriction mapping of
 CC the clone and multiple cloning steps. The method of the invention is fast
 CC and efficiently generates nucleotide construct, and reduces the time
 CC required for making a knock-out vector. Isolating an individual genomic
 CC clone or mapping the restriction sites within the clone is not needed for
 CC the method. The method is ligation independent cloning. The present
 CC nucleic acid sequence represents the PDG2 plasmid vector sequence of the
 CC invention
 XX
 SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 6; Length 4768;
 Best Local Similarity 100.0%; Pred. No. 4.2e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTTAACACGTCAGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
 DB 1 GTTAACACGTCAGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
 QY 61 TTCTTAATACATTCAAATATGTATCGCTCATGAGACATAACCTGATAATGCTTCA 120
 DB 61 TTCTTAATACATTCAAATATGTATCGCTCATGAGACATAACCTGATAATGCTTCA 120
 QY 121 ATAAATTTGAAAAGGAAGAGTATGATGATTCACATTTCCGTCGCGCTTATTCCTT 180
 DB 121 ATAAATTTGAAAAGGAAGAGTATGATGATTCACATTTCCGTCGCGCTTATTCCTT 180
 QY 181 TTTTGGCGCATTTTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 240

QY 421 ACACATATCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAAGCATCTTACGGA 480
 Db 421 ACACATATCTCAGATGACTTGGTTGAGTACTACCAAGTACAGAAAAGCATCTTACGGA 480
 QY 481 TGGCATGACAGTAAGAGAAT 500
 Db 481 TGGCATGACAGTAAGAGAAT 500
 RESULT 10
 AAS05243
 ID AAS05243 standard; DNA; 4768 BP.
 XX
 AC AAS05243;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Plasmid vector pDG2 used as a construct for TRP genes.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES; pDG2;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; cyclic; circular; ds.
 XX
 OS Synthetic.
 XX
 PN WO200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 XX
 PF 26-OCT-2000; 2000WO-US029382.
 XX
 PR 26-OCT-1999; 99US-0161488P.
 XX
 PA (DELT-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen XD;
 XX
 DR WPI; 2001-300473/31.
 XX
 PT Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP.
 XX
 PS Disclosure; Fig 2B; 106pp; English.
 XX
 CC The present sequence for plasmid vector pDG2 is used as a construct for
 CC genes encoding trinucleotide repeat proteins (TRP) such as gene T243 to
 CC produce disruption in the DNA. The invention describes methods of
 CC producing embryonic stem (ES) cells comprising a heterozygous disruption
 CC in a target DNA sequence (preferably T243) encoding a TRP and of
 CC producing a knockout mouse comprising a homozygous disruption in a gene
 CC encoding TRP, where the disruption inhibits the production of the wild
 CC type TRP. The invention also relates to identifying agents capable of
 CC affecting a phenotype of a knockout mouse. Also described are methods of
 CC determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and the
 CC cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse
 XX
 SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;
 Query Match 100.0%; Score 500; DB 4; Length 4768;
 Best Local Similarity 100.0%; Pred. NO. 4.2e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACTACGTGAGTGGACATTTTCGGGGAATGTCGGCGGAAACCCCTATTGTTTATT 60
 Db 1 GTTAACTACGTGAGTGGACATTTTCGGGGAATGTCGGCGGAAACCCCTATTGTTTATT 60
 QY 61 TTTCTAAATACATTTCAAATATGATATCCGCTCATGAGACAATTAACCTGATAAATGCTTCA 120
 Db 61 TTTCTAAATACATTTCAAATATGATATCCGCTCATGAGACAATTAACCTGATAAATGCTTCA 120
 QY 121 ATAATATTCAAAAAGGAGATGATGAGTATTCAACATTTCCGTCGCGCCCTTATTCCCTT 180
 Db 121 ATAATATTCAAAAAGGAGATGATGAGTATTCAACATTTCCGTCGCGCCCTTATTCCCTT 180
 QY 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
 Db 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
 QY 241 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTCATCGAACTGGATCTCAACAGCGGTAA 300
 Db 241 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTCATCGAACTGGATCTCAACAGCGGTAA 300
 QY 301 GATCCTTGAGAGTTTTCGCCCCGAGAGAGTTCCTCAATGATGAGACATTTTAAAGTTCT 360
 Db 301 GATCCTTGAGAGTTTTCGCCCCGAGAGAGTTCCTCAATGATGAGACATTTTAAAGTTCT 360
 QY 361 GCTATGTCGCGCGGTATTATCCCTGTTTGAACCGCGGCAAGAGCAACTCGGTCGCGCAT 420
 Db 361 GCTATGTCGCGCGGTATTATCCCTGTTTGAACCGCGGCAAGAGCAACTCGGTCGCGCAT 420
 QY 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACAGTACAGAAAAGCATCTTACGGA 480
 Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTCACAGTACAGAAAAGCATCTTACGGA 480
 QY 481 TGGCATGACAGTAAGAGAAT 500
 Db 481 TGGCATGACAGTAAGAGAAT 500
 RESULT 11
 ABL42019
 ID ABL42019 standard; DNA; 4768 BP.
 XX
 AC ABL42019;
 XX
 DT 11-JUN-2002 (first entry)
 XX
 DE Nucleotide sequence of vector pDG2.
 XX
 KW pDG2; transgenic animal; matrix metalloproteinase-23 gene; MMP-23 gene;
 KW SS.
 XX
 OS Synthetic.
 XX
 PN US2002023275-A1.
 XX
 PD 21-FEB-2002.
 XX
 PP 17-MAY-2001; 2001US-00861077.
 XX
 PR 17-MAY-2000; 2000US-0204972P.
 PR 29-JUN-2000; 2000US-0215394P.
 XX
 PA (LEVI/) LEVITEN M W.
 XX
 PI Leviten MW;
 XX
 DR WPI; 2002-255684/30.
 XX
 PT Non-human transgenic animal useful as a model for disease and for
 PT identifying agents that modulate gene expression and gene function,
 PT comprises a disruption in the matrix metalloproteinase-23 gene.
 XX
 PS Example 1; Fig 2B; 38pp; English.
 XX

transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or mammal) by introducing (ii) by cell fusion, lipid-mediated transfection, by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer into an embryonic cell, preferably a stem cell or an embryo. (iii) comprises a heterologous nucleic acid that encodes a therapeutic product which is useful for making a library of ACes comprising random portions of a genome. ACC44612 to ACC44732 and ABP96650 to ABP96657 represent sequences used in the exemplification of the present invention

Sequence 4223 BP; 1143 A; 966 C; 969 G; 1145 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 7; Length 4223;
Best Local Similarity 100.0%; Pred. No. 4e-124;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACTACGTCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
DB GTTAACTACGTCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATT 3868

QY 61 TTTCTAATATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 120
DB TTTCTAATATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 3808

QY 121 ATAAATATTGAAAGAGAGATGAGTATTCACATTTCCGTGTGCGCCCTTATTCCTT 180
DB ATAAATATTGAAAGAGAGATGAGTATTCACATTTCCGTGTGCGCCCTTATTCCTT 3748

QY 181 TTTTCGGCATTTCGCTTCTGTTTTCGCTCACCCAGAAACGGTGTGAAAGTAAAGA 240
DB TTTTCGGCATTTCGCTTCTGTTTTCGCTCACCCAGAAACGGTGTGAAAGTAAAGA 3688

QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATTCATCGAATGATCTCAACGCGGTAA 300
DB TGCTGAAGATCAGTTGGGTGACGAGTGGGTATTCATCGAATGATCTCAACGCGGTAA 3628

QY 301 GATCCTTGAGAGTTTTCGCGCGAAGAACGTTCTCCAAATGATGACACTTTTAAAGTTCT 360
DB GATCCTTGAGAGTTTTCGCGCGAAGAACGTTCTCCAAATGATGACACTTTTAAAGTTCT 3568

QY 361 GCTATGTGCGCGGTATTATCCGTTGTCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
DB GCTATGTGCGCGGTATTATCCGTTGTCGCGGCAAGAGCAACTCGGTGCGCGCAT 3508

QY 421 ACACATTTCTCAGAATGACTTGGTTGAGTACTCACTCAGTACAGAAAGCATCTTACGGA 480
DB ACACATTTCTCAGAATGACTTGGTTGAGTACTCACTCAGTACAGAAAGCATCTTACGGA 3448

QY 481 TGGCATGACATGAGAGAT 500
DB TGGCATGACATGAGAGAT 3428

RESULT 9
ACC44714
ID ACC44714 standard; DNA; 4223 BP.
XX
AC ACC44714;
XX
DT 29-MAY-2003 (first entry)
XX
DE Plasmid pLIT38attcBBSRpolyA10 nucleotide sequence SEQ ID NO:11.
XX
KW Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
att site; integrase; recombinase; ACes; gene therapy; transgenic animal;
platform artificial chromosome expression system; gene; ds.
XX
OS Synthetic.
XX
PN WO200297059-A2.
XX
PD 05-DEC-2002.
XX

30-MAY-2002; 2002WO-US017452.
30-MAY-2001; 2001US-0294758P.
21-MAR-2002; 2002US-0366891P.
(CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
Stewart S, Shellard J;
WPI; 2003-140461/13.
Novel eukaryotic chromosome comprising one or many att sites which permits site-directed integration in the presence of lambda-integrase, useful for site-specific recombination-directed integration of DNA of interest.
Example 10; Page 241-242; 272pp; English.
The present invention describes a eukaryotic chromosome (i) comprising one or several att sites, where an att site is heterologous to the chromosome, and permits site-directed integration in the presence of lambda-integrase. Also described: (1) a platform artificial chromosome expression system (Acce) (ii) comprising several sites that participate in recombination catalysed recombination; and (2) a method (M1) for introducing a heterologous nucleic acid into a platform artificial chromosome. (i) can be used in gene therapy. (M1) is useful for introducing a heterologous nucleic acid molecule into a platform artificial chromosome, preferably an ACes. (ii) is useful for producing a transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or mammal) by introducing (ii) by cell fusion, lipid-mediated transfection, by a carrier system, microinjection, microcell fusion, electroporation, microprojectile bombardment or direct DNA transfer into an embryonic cell, preferably a stem cell or an embryo. (iii) comprises a heterologous nucleic acid that encodes a therapeutic product which is useful for making a library of ACes comprising random portions of a genome. ACC44612 to ACC44732 and ABP96650 to ABP96657 represent sequences used in the exemplification of the present invention

Sequence 4223 BP; 1185 A; 931 C; 1004 G; 1103 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 7; Length 4223;
Best Local Similarity 100.0%; Pred. No. 4e-124;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACTACGTCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATT 60
DB GTTAACTACGTCAGGTGGCAGCTTTTCGGGGAAATGTGCGGAAACCCCTATTGTTTATT 60

QY 61 TTTCTAATATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 120
DB TTTCTAATATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 120

QY 121 ATAAATATTGAAAGAGAGATGAGTATTCACATTTCCGTGTGCGCCCTTATTCCTT 180
DB ATAAATATTGAAAGAGAGATGAGTATTCACATTTCCGTGTGCGCCCTTATTCCTT 180

QY 181 TTTTCGGCATTTCGCTTCTGTTTTCGCTCACCCAGAAACGGTGTGAAAGTAAAGA 240
DB TTTTCGGCATTTCGCTTCTGTTTTCGCTCACCCAGAAACGGTGTGAAAGTAAAGA 240

QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATTCATCGAATGATCTCAACGCGGTAA 300
DB TGCTGAAGATCAGTTGGGTGACGAGTGGGTATTCATCGAATGATCTCAACGCGGTAA 300

QY 301 GATCCTTGAGAGTTTTCGCGCGAAGAACGTTCTCCAAATGATGACACTTTTAAAGTTCT 360
DB GATCCTTGAGAGTTTTCGCGCGAAGAACGTTCTCCAAATGATGACACTTTTAAAGTTCT 360

QY 361 GCTATGTGCGCGGTATTATCCGTTGTCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
DB GCTATGTGCGCGGTATTATCCGTTGTCGCGGCAAGAGCAACTCGGTGCGCGCAT 420

autoimmune disease; transplant rejection; systemic lupus erythematosus;
type I diabetes; rheumatoid arthritis; myasthenia gravis;
multiple sclerosis; AIDS; acquired immunodeficiency syndrome;
chronic immunosuppression.

Unidentified.

WO200197982-A2.

22-NOV-2001.

18-MAY-2001; 2001WO-US016125.

18-MAY-2000; 2000US-00573797.

(NOVS) NOVARTIS AG.

(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

(USSH) US DEPT HEALTH & HUMAN SERVICES.

Neville DM, Thompson JT, Hu H, Woo J, Ma S, Hexham JM, Digan ME;

WPI; 2002-121980/16.

New anti-T cell immunotoxin fusion protein comprising a truncated
diphtheria toxin moiety, a connector, and one single chain Fv of the
variable region of a UCHT1 antibody, useful for treating e.g. autoimmune
diseases.

Disclosure; Page 255-256; 307pp; English.

The invention relates to an anti-T cell immunotoxin fusion protein,
comprising from the amino terminus, a truncated diphtheria toxin moiety,
a connector, and one single chain Fv of the variable region of a UCHT1
antibody. The single chain Fv comprises VL, L, or VH, L is a Gly-Ser
linker, and VL and VH are the variable light and heavy domains of the
anti-CD3 antibody UCHL1. Also included are a nucleic acid encoding the
novel immunotoxin fusion protein, a vector comprising the nucleic acid
and a cell comprising the nucleic acid. The immunotoxin may also be a
pseudomonas endotoxin A (ETA). The immunotoxins are useful for treating T
cell leukaemias or lymphomas, graft-versus-host diseases, and autoimmune
diseases by inducing immune tolerance. The immunotoxin fusion proteins
may be used in vivo to systemically reduce populations of T cells, or ex
vivo to effect T-cell depletion from a treated cell population. The
fusion proteins can be administered to a subject who is or will be a
recipient of an allotransplant to prevent or reduce T-cell mediated acute
or chronic transplant rejection of the transplanted allogeneic cells,
tissue or organ in the subject as well as treat other T-cell mediated
diseases such as systemic lupus erythematosus, type I diabetes,
rheumatoid arthritis, myasthenia gravis, multiple sclerosis, infectious
diseases of the immune system (e.g. AIDS (acquired immunodeficiency
syndrome)) and chronic immunosuppression. The present sequence is a
plasmid, pCE96, used as an intermediate cloning vector to construct the
immunotoxin fusion proteins of the invention

Sequence 3476 BP; 911 A; 925 C; 863 G; 777 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 6; Length 3476;
Best Local Similarity 100.0%; Pred. No. 3.8e-124; Mismatches 0; Gaps 0;
Matches 500; Conservative 0; Indels 0; Indels 0; Gaps 0;

1 G T T A C T A C G T C A G T G G C A C T T T C G G G A A A T G C G C G A A C C C C T A T T G T T A T T 60
3476 G T T A C T A C G T C A G T G G C A C T T T C G G G A A A T G C G C G A A C C C C T A T T G T T A T T 3417
61 T T T C T A A T A C A T T C A A T A T A T C C G T C A T G A C A A T A A C C C T A A T G C T T C A 120
3416 T T T C T A A A T A C A T T C A A T A T A T C C G T C A T G A C A A T A A C C C T A A T G C T T C A 3357
121 A T A T A T T G A A A A G A A G A T A G A T T C A C A T T T C C G T G C G C C C T A T T C C C T T 180
3356 A T A T A T T G A A A A G A A G A T A G A T T C A C A T T T C C G T G C G C C C T A T T C C C T T 3297
181 T T T G C G C A T T T G C C T T C C T G T T T T G C T C A C C A G A A C G C T G G T G A A A G A A A G A 240

Db 3296 T T T T G C G C A T T T T G C C T T C C T G T T T T T G C T C A C C C A A A C G C T G T G A A G T A A A A G A 3237
Qy 241 T G C T A A G A T C A G T T G G G T G C A C G A G T G G T T A C A T C G A A C T G A T C T C A C A C G G T A A 300
Db 3236 T G C T A A G A T C A G T T G G G T G C A C G A G T G G T T A C A T C G A A C T G A T C T C A C A C G G T A A 3177
Qy 301 G A T C C T T G A G A G T T T T C C C C C G A A G A A C G T T C C A A T A T G A T G A G A C A C T T T T A A G T T C T 360
Db 3176 G A T C C T T G A G A G T T T T C G C C C G A A G A A C G T T C C A A T G A T G A G A C A C T T T T A A G T T C T 3117
Qy 361 G C T A T G T G C G C G G T A T T A T C C G T G T T G A C G C G G G C A A G A C A A C T C G G T C G C C G C A T 420
Db 3116 G C T A T G T G C G C G G T A T T A T C C G T G T T G A C G C G G G C A A G A C A A C T C G G T C G C C G C A T 3057
Qy 421 A C A C T A T T T C T C A G A A T G A C T T G G T T G A T A C C A G T C A C A G A A A G A A G C A T C T T A C G G A 480
Db 3056 A C A C T A T T T C T C A G A A T G A C T T G G T T G A T A C C A G T C A C A G A A A A G A A G C A T C T T A C G G A 2997
Qy 481 T G G C A T G A C A G T A A G A G A A T 500
Db 2996 T G G C A T G A C A G T A A G A A T 2977

RESULT 8

ACC44724/c

ID ACC44724 standard; DNA; 4223 BP.

XX AC ACC44724;

XX DT 29-MAY-2003 (first entry)

XX DE Plasmid pLIT3attB-BSDpolyA2 nucleotide sequence SEQ ID NO:121.

XX KW Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
att site; integrase; recombinase; ACes; Gene therapy; transgenic animal;
XX KW platform artificial chromosome expression system; Gene; ds.

XX OS Synthetic.

XX PN WO200297059-A2.

XX PD 05-DEC-2002.

XX PF 30-MAY-2002; 2002WO-US017452.

XX PR 30-MAY-2001; 2001US-0294758P.

XX PR 21-MAR-2002; 2002US-0366891P.

XX (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.

XX PA Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
XX PI Stewart S, Shellard J;

XX WPI; 2003-140461/13.

XX Novel eukaryotic chromosome comprising one or many att sites which
permits site-directed integration in the presence of lambda-integrase,
useful for site-specific recombination-directed integration of DNA of
interest.

XX Example 10; Page 258-259; 272pp; English.

XX The present invention describes a eukaryotic chromosome (I) comprising
one or several att sites, where an att site is heterologous to the
chromosome, and permits site-directed integration in the presence of
lambda-integrase. Also described: (I) a platform artificial chromosome
expression system (AcEs) (II) comprising several sites that participate
in recombinase catalysed recombination; and (2) a method (M1) for
introducing a heterologous nucleic acid into a platform artificial
chromosome. (I) can be used in gene therapy. (M1) is useful for
introducing a heterologous nucleic acid molecule into a platform
artificial chromosome, preferably an ACes. (II) is useful for producing a

481 TCGTGAAGATCAGTTGGTGCACGAGTGGGTTATCATCGAACTGGATCTCAACACGGGTAA 540
301 GATCCTTGAGAGTTTCCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 360
541 GATCCTTGAGAGTTTCCGCCCGGAGAACGTTCTCCAAATGATGAGCACTTTTAAAGTTCT 600
361 GCTATGTGGCGCGGTATATCCCGTGTGACGCGCGGAGAGCAACTCGTGCAGCAT 420
601 GCTATGTGGCGCGGTATATCCCGTGTGACGCGCGGAGAGCAACTCGTGCAGCAT 660
421 ACACATTTCTCAGAAATGATCTGGTTGAGTACTACACAGTACACAGAAAAGCATCTTACGGA 480
661 ACACATTTCTCAGAAATGATCTGGTTGAGTACTACACAGTACACAGAAAAGCATCTTACGGA 720
481 TGGCATGACAGTAAGAGAAAT 500
721 TGGCATGACAGTAAGAGAAAT 740
RESULT 6
AAV33299/c
ID AAV33299 standard; DNA; 3476 BP.
XX AAV33299;
AC AAV33299;
XX
XX
17-OCT-2003 (revised)
DT 21-DEC-1998 (first entry)
XX
D3 E.coli-Corynebacterium shuttle vector yCE96.
XX
XX Vector; yCE96; plasmid pNG2; Litmus p29; toxin; immunotoxin;
KW diphtheria toxin; Pseudomonas exotoxin A; ss.
XX
OS Escherichia coli.
OS Corynebacterium sp.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT misc_feature 1. 373
FT /tag= a
FT /note= "from vector LITMUS 29, contains polycloning
FT linker sites"
FT 374. 2152
FT /tag= b
FT /function= "origin sequences from plasmid pNG2"
FT 2153. 3476
FT /tag= c
FT /function= "ampicillin resistance marker"
XX
PN WO9839425-A2.
XX
PD 11-SEP-1998.
XX
XX 05-MAR-1998; 98WO-US004155.
XX
XX 05-MAR-1997; 97US-0037196P.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Neville DM;
XX
XX WPI; 1998-495837/42.
XX
XX New Escherichia coli and Corynebacterium shuttle vector - useful for the
XX expression of mutant toxins in resistant Pichia pastoris, Chinese hamster
XX ovary and insect cells having a mutated translation elongation factor 2.
XX
XX Example 1; Page 24-25; 34pp; English.
XX
XX This is the nucleotide sequence of a new E. coli/Corynebacterium shuttle
XX vector, yCE96, constructed using the origin of plasmid pNG2 (from
XX Escherichia coli JM109) and the antibiotic resistance marker and multiple

CC cloning sites of the vector Litmus p29. The vector is only 3.4 kb in size
CC and can transform both E. coli and Corynebacterium ulcerans and, thus,
CC can be used to produce toxins and mutant toxins. In addition, the
CC invention provides a mutant Pichia pastoris, a method for producing this
CC mutant and a method of expressing engineered toxin mutants and toxin
CC fusion proteins in the mutant P. pastoris. The invention further provides
CC a mutant CHO cell, a mutant insect cell, a method for producing these
CC mutants and a method of expressing engineered toxin mutants and toxin
CC fusion proteins in the mutant cells. Methods are claimed for expressing a
CC diphtheria toxin or a Pseudomonas exotoxin A toxin in the mutant cells.
CC The invention results in higher fusion protein secretion than prior art
CC methods, thus eliminating the need for refolding, and providing
CC glycosylated immunotoxins that are useful for in vivo clinical
CC application. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ Sequence 3476 BP; 911 A; 925 C; 863 G; 777 T; 0 U; 0 Other;
Query Match 100.0%; Score 500; DB 2; Length 3476;
Best Local Similarity 100.0%; Pred. No. 3.8e-124;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACACTGCTCAGGTGGCACTTTTCGGGGAATGTCGGCGAAGCCCTATTGTTTATT 60
DB 3476 GTTAACACTGCTCAGGTGGCACTTTTCGGGGAATGTCGGCGAAGCCCTATTGTTTATT 3417
QY 61 TTTCTAAATACATTCAAATATGATATCGCTCATGAGACAATAACCTCATAAATGCTTCA 120
DB 3416 TTTCTAAATACATTCAAATATGATATCGCTCATGAGACAATAACCTCATAAATGCTTCA 3357
QY 121 ATAAATATTGAAAAAGAGAGATGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCTT 180
DB 3356 ATAAATATTGAAAAAGAGAGATGATGAGTATTCACATTTCCGTGTCGCCCTTATTCCTT 3297
QY 181 TTTTGGCGCATTTTGCCTTCCTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 240
DB 3296 TTTTGGCGCATTTTGCCTTCCTCTGTTTTCCTCACCAGAAACGCTGGTGAAGTAAAGA 3237
QY 241 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAATCTGATCTCAACAGCGGTAA 300
DB 3236 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAATCTGATCTCAACAGCGGTAA 3177
QY 301 GATCCTTGAGAGTTTTCGCCCGGAGAAAGCTTCTCCAATGATGACACTTTTAAAGTTCT 360
DB 3176 GATCCTTGAGAGTTTTCGCCCGGAGAAAGCTTCTCCAATGATGACACTTTTAAAGTTCT 3117
QY 361 GCTATGTGGCGCGGTATTATCCCGTGTGACGCGCGGCAAGCAACTCGTTCGCCGAT 420
DB 3116 GCTATGTGGCGCGGTATTATCCCGTGTGACGCGCGGCAAGCAACTCGTTCGCCGAT 3057
QY 421 ACACATTTCTCAGAAATGATCTGGTTGAGTACTCACCAGTACAGAAAAGCATCTTACGGA 480
DB 3056 ACACATTTCTCAGAAATGATCTGGTTGAGTACTCACCAGTACAGAAAAGCATCTTACGGA 2997
QY 481 TGGCATGACAGTAAGAGAAAT 500
DB 2996 TGGCATGACAGTAAGAGAAAT 2977
RESULT 7
ABK13447/c
ID ABK13447 standard; DNA; 3476 BP.
XX
XX ABK13447;
XX
XX 23-APR-2002 (first entry)
XX
XX Plasmid pCE96 for mutation of immunotoxins.
XX
XX pCE96; ds; immunotoxin; pseudomonas exotoxin A; ETA; diphtheria toxin;
XX DT; cytostatic; immunosuppressive; immunostimulant; antidiabetic;
XX anti-rheumatic; antiarthritic; anti-HIV; anti-inflammatory;
XX anti-T cell immunotoxin fusion protein; antibody; UCHL1; Fv; CD3;
XX T cell leukaemia; lymphoma; graft-versus-host disease;

XX The present invention describes a eukaryotic chromosome (I) comprising
 CC one or several att sites, where an att site is heterologous to the
 CC chromosome, and permits site-directed integration in the presence of
 CC lambda-integrase. Also described: (1) a platform artificial chromosome
 CC expression system (Acet) (II) comprising several sites that participate
 CC in recombinase catalysed recombination; and (2) a method (M1) for
 CC introducing a heterologous nucleic acid into a platform artificial
 CC chromosome. (I) can be used in gene therapy. (M1) is useful for
 CC introducing a heterologous nucleic acid molecule into a platform
 CC artificial chromosome, preferably an Acet. (II) is useful for producing a
 CC transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
 CC mammal) by introducing (II) by cell fusion, lipid-mediated transfection,
 CC by a carrier system, microinjection, microcell fusion, electroporation,
 CC microprojectile bombardment or direct DNA transfer into an embryonic
 CC cell, preferably a stem cell or an embryo. (II) comprises a heterologous
 CC nucleic acid that encodes a therapeutic product which is useful for
 CC making a library of Acet comprising random portions of a genome. ACC44612
 CC to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
 CC exemplification of the present invention
 XX
 XX Sequence 3438 BP; 849 A; 854 C; 902 G; 833 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 7; Length 3438;
 Best Local Similarity 100.0%; Pred. No. 3.8e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTTACTAGCTCAGTGGGCACTTTTCGGGGAATGTGCGGAAATGTCGCCATTTTGTATT 60
 DB 241 GTTAACTAGCTCAGTGGGCACTTTTCGGGGAATGTGCGGAAATGTCGCCATTTTGTATT 300
 QY 61 TTTCTAAATACATTTCAATATATGATTCGGTTCATGAGACAATTAACCTGATAAATGCTTCA 120
 DB 301 TTTCTAAATACATTTCAATATATGATTCGGTTCATGAGACAATTAACCTGATAAATGCTTCA 360
 QY 121 ATATATTTGAAAAGGAGATGAGTATTCATTCGATTCGATTCGATTCGATTCGATTCGATTC 180
 DB 361 ATATATTTGAAAAGGAGATGAGTATTCATTCGATTCGATTCGATTCGATTCGATTCGATTC 420
 QY 181 TTTTGGCGCATTTTGGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 240
 DB 421 TTTTGGCGCATTTTGGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 480
 QY 241 TGTGTAAGATCATGTTGGGTGCGAGTGGGTACATCGAATCGAATCGAATCGAATCGAATCG 300
 DB 481 TGTGTAAGATCATGTTGGGTGCGAGTGGGTACATCGAATCGAATCGAATCGAATCGAATCG 540
 QY 301 GATCCTTGAGAGTTTTCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 DB 541 GATCCTTGAGAGTTTTCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 361 GCTATGTGGCGGATTTATTCGGTGTGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 601 GCTATGTGGCGGATTTATTCGGTGTGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 421 ACATATTTCTCAGATGACTTGGTTCAGTACTCAGTACTCAGTACTCAGTACTCAGTACTCAG 480
 DB 661 ACATATTTCTCAGATGACTTGGTTCAGTACTCAGTACTCAGTACTCAGTACTCAGTACTCAG 720
 QY 481 TGGCATGACAGTAAGAGAT 500
 DB 721 TGGCATGACAGTAAGAGAT 740

RESULT 5
 ABT16608
 ID ABT16608 standard; DNA; 3438 BP.
 XX
 AC ABT16608;
 XX
 DT 03-APR-2003 (first entry)
 XX
 DE Artificial plant chromosome related plasmid DNA SEQ ID No 19.

XX Plant artificial chromosome; PAC; transgenic plant; vaccine;
 KW blood factor; herbicide; stress; agronomical; nutrient quality;
 KW bacterial artificial chromosome; BAC; yeast artificial chromosome; YAC;
 KW ds.
 XX Unidentified.
 XX WO200296923-A1.
 XX 05-DEC-2002.
 XX 30-MAY-2002; 2002WO-US017451.
 XX 30-MAY-2001; 2001US-0294687P.
 XX 04-JUN-2001; 2001US-0296329P.
 XX (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
 XX (AGRI-) AGRISOMA INC.
 XX Perez C, Fabijanski SF, Perkins E;
 XX WPI; 2003-140436/13.
 XX Producing artificial chromosome by introducing a nucleic acid into plant
 PT cell, selecting artificial chromosome that has one or more repeat regions
 PT with equivalent amounts of euchromatic and heterochromatic nucleic acids.
 XX Example 5; Page 247-248; 269pp; English.
 XX The invention relates to a novel method for producing plant artificial
 CC chromosomes. The invention also relates to methods for targeting, methods
 CC insertion of heterologous DNA into plant artificial chromosomes, methods
 CC for delivery of plant chromosomes to selected cells and tissues. The
 CC isolated plant artificial chromosome (PAC) is useful for producing a
 CC transgenic plant, which involves introducing the PAC into a plant cell.
 CC The PAC comprises a heterologous nucleic acid encoding a gene product
 CC such as enzymes, antisense RNA, tRNA, rDNA, structural proteins, marker
 CC proteins, ligands, receptors, ribozymes, therapeutic proteins, and
 CC biopharmaceutical proteins, vaccines, blood factors, antigens, hormones,
 CC cytokines, growth factors, antibodies, or a product that provides for
 CC resistance to diseases, insects, herbicides, or stress in a plant. The
 CC heterologous nucleic acid optionally encodes a product that provides an
 CC agronomically important trait in the plant, e.g. a product that alters
 CC nutrient use and/or improves the nutrient quality of the plant. The
 CC heterologous nucleic acid is contained within a bacterial artificial
 CC chromosome (BAC) or a yeast artificial chromosome (YAC). This
 CC method of the invention

Sequence 3438 BP; 849 A; 854 C; 902 G; 833 T; 0 U; 0 Other;
 Query Match 100.0%; Score 500; DB 7; Length 3438;
 Best Local Similarity 100.0%; Pred. No. 3.8e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTTAACTAGCTCAGTGGGCACTTTTCGGGGAATGTGCGGAAATGTCGCCATTTTGTATT 60
 DB 241 GTTAACTAGCTCAGTGGGCACTTTTCGGGGAATGTGCGGAAATGTCGCCATTTTGTATT 300
 QY 61 TTTCTAAATACATTTCAATATATGATTCGGTTCATGAGACAATTAACCTGATAAATGCTTCA 120
 DB 301 TTTCTAAATACATTTCAATATATGATTCGGTTCATGAGACAATTAACCTGATAAATGCTTCA 360
 QY 121 ATATATTTGAAAAGGAGATGAGTATTCATTCGATTCGATTCGATTCGATTCGATTCGATTC 180
 DB 361 ATATATTTGAAAAGGAGATGAGTATTCATTCGATTCGATTCGATTCGATTCGATTCGATTC 420
 QY 181 TTTTGGCGCATTTTGGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 240
 DB 421 TTTTGGCGCATTTTGGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 480
 QY 241 TGCTGAAGATCAGTGGGTGCGAGTGGGTACATCGAATCGAATCGAATCGAATCGAATCG 300

```

Db      361 GCTATGTGCGCGGTATATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGCCAT 420
Qy      421 ACACATTCTCAGAAAGCATTGGTTGAGTACTACCAAGTACACAGAAAAGCATCTTACGGA 480
Db      421 ACACATTCTCAGAAAGCATTGGTTGAGTACTACCAAGTACACAGAAAAGCATCTTACGGA 480
Qy      481 TGGCATGACAGTAAGAGAAAT 500
Db      481 TGGCATGACAGTAAGAGAAAT 500

RESULT 3
AA14902
ID AA14902 standard; DNA; 3419 BP.
XX AC
XX AA14902;
XX
XX 08-AUG-2000 (first entry)
XX Nucleotide sequence of plasmid pLITMUS-IFN-gamma.
XX Interferon-gamma; RNA quantification; pLITMUS-IFN-gamma; ss.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX misc_feature 2530..3027
XX /*tag= a
XX /*note= "interferon-gamma specific sequence"
XX promoter 3199..3216
XX /*tag= b
XX /*note= "T7 promoter"
XX
XX WO200020629-A2.
XX
XX PD 13-APR-2000.
XX
XX PF 04-OCT-1999; 99WO-CA000917.
XX
XX PR 05-OCT-1999; 98US-0103153P.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX
XX PI Parrington M, Caterini JE, Klein MH;
XX WPI; 2000-328942/28.
XX
XX PT Accurate method of quantifying RNA in a sample by comparing labelled
XX oligonucleotide bound to target RNA with oligonucleotide bound to an
XX internal standard.
XX
XX PS Example 5; Fig 2A-B; 32pp; English.
XX
XX CC The present sequence represents a plasmid which is used to generate an
XX interferon-gamma RNA standard for use to exemplify the method of the
XX invention. The specification describes a method of determining the
XX quantity of a target RNA in a tissue sample. The method comprises reverse
XX transcribing target RNA from the tissue and then amplifying the cDNA,
XX binding a labelled sequence corresponding to an internal sequence
XX complementary to one of the strands of the target RNA product, and
XX determining the amount of labelled RNA bound to the amplified target DNA.
XX The method is used to quantify more accurately the amount of target RNA
XX in a sample
XX
XX SQ Sequence 3419 BP; 892 A; 806 C; 840 G; 881 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 500; DB 3; Length 3419;
XX Best Local Similarity 100.0%; Pred. No. 3.8e-124;
XX Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy      1 GTTAAGTACGTCAGTGGCAGCTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATT 60

```

```

Db      61 GTTAAGTACGTCAGTGGCAGCTTTTCGGGAAATGTGCGGGAACCCCTATTGTTTATT 60
Qy      61 TTTCTAAATACATTCAAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 120
Db      61 TTTCTAAATACATTCAAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCA 120
Qy      121 ATAATATTGAAAAGGAAGAGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTT 180
Db      121 ATAATATTGAAAAGGAAGAGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTT 180
Qy      181 TTTTGGGCAATTTTGCCTTCTCTGTTTCTCACCAGAAACCGTGGTGAAGTAAAGA 240
Db      181 TTTTGGGCAATTTTGCCTTCTCTGTTTCTCACCAGAAACCGTGGTGAAGTAAAGA 240
Qy      241 TGTGAAGATCAGTTGGGTGACGAGTGGTTCATCGAATCTGGATCTCAACAGCGGTAA 300
Db      241 TGTGAAGATCAGTTGGGTGACGAGTGGTTCATCGAATCTGGATCTCAACAGCGGTAA 300
Qy      301 GATCCTTGAGAGTTTTCGCCCGCAAGAGAGTCTCAATGATGAGACATTTTAAAGTTCT 360
Db      301 GATCCTTGAGAGTTTTCGCCCGCAAGAGAGTCTCTCAATGATGAGACATTTTAAAGTTCT 360
Qy      361 GCTATGTGCGCGGTATATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db      361 GCTATGTGCGCGGTATATATCCCGTGTGACGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Qy      421 ACACATTCTCAGAAAGCATTGGTTGAGTACTACCAAGTACACAGAAAAGCATCTTACGGA 480
Db      421 ACACATTCTCAGAAAGCATTGGTTGAGTACTACCAAGTACACAGAAAAGCATCTTACGGA 480
Qy      481 TGGCATGACAGTAAGAGAAAT 500
Db      481 TGGCATGACAGTAAGAGAAAT 500

RESULT 4
ACCA4695
ID ACC44695 standard; DNA; 3438 BP.
XX AC
XX ACC44695;
XX
XX 29-MAY-2003 (first entry)
XX
XX DE Plasmid pLITattBZeo nucleotide sequence SEQ ID NO:91.
XX
XX KW Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
XX att site; integrase; recombinase; Aces; gene therapy; transgenic animal;
XX platform artificial chromosome expression system; gene; ds.
XX
XX OS Synthetic.
XX
XX PN WO200297059-A2.
XX
XX PD 05-DEC-2002.
XX
XX PF 30-MAY-2002; 2002WO-US017452.
XX
XX PR 30-MAY-2001; 2001US-0294758P.
XX
XX PR 21-MAR-2002; 2002US-0366891P.
XX
XX PA (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX
XX PI Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
XX Stewart S, Shellard J;
XX WPI; 2003-140461/13.
XX
XX PT Novel eukaryotic chromosome comprising one or many att sites which
XX permits site-directed integration in the presence of lambda-integrase,
XX useful for site-specific recombination-directed integration of DNA of
XX interest.
XX
XX PS Example 7; Page 217-218; 272pp; English.

```

CC chromosome. (I) can be used in gene therapy. (M1) is useful for
 CC introducing a heterologous nucleic acid molecule into a platform
 CC artificial chromosome, preferably an ACes. (II) is useful for producing a
 CC transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
 CC mammal) by introducing (III) by cell fusion, lipid-mediated transfection,
 CC by a carrier system, microinjection, microcell fusion, electroporation,
 CC microprojectile bombardment or direct DNA transfer into an embryonic
 CC cell, preferably a stem cell or an embryo. (II) comprises a heterologous
 CC nucleic acid that encodes a therapeutic product which is useful for
 CC making a library of ACes comprising random portions of a genome. ACC44612
 CC to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
 CC exemplification of the present invention

XX Sequence 2814 BP; 733 A; 679 C; 714 G; 688 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 7; Length 2814;
 Best Local Similarity 100.0%; Pred. No. 3.6e-124; Indels 0; Gaps 0;
 Matches 500; Conservative 0; Mismatches 0;

QY 1 GTTAACTACGTGAGTGGGACATTTTCGGGGAATGTGGCGGAACCCCTATTGTTATT 60
 Db 1 GTTAACTACGTGAGTGGGACATTTTCGGGGAATGTGGCGGAACCCCTATTGTTATT 60
 QY 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGACATTAACCTGATTAATCTTCA 120
 Db 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGACATTAACCTGATTAATCTTCA 120
 QY 121 ATAATATTGAAAAGGAAGATGATGATTCAACATTTCCGTCGCGCCCTATTCCCTT 180
 Db 121 ATAATATTGAAAAGGAAGATGATGATTCAACATTTCCGTCGCGCCCTATTCCCTT 180
 QY 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGA 240
 Db 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGA 240
 QY 241 TGTGTAAGATCAGTGGGTGACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA 300
 Db 241 TGTGTAAGATCAGTGGGTGACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA 300
 QY 301 GATCCTTGAGAGTTTTCGCCCGAAGACGTTCTCCATGATGAGACACTTTTAAAGTCT 360
 Db 301 GATCCTTGAGAGTTTTCGCCCGAAGACGTTCTCCATGATGAGACACTTTTAAAGTCT 360
 QY 361 GCTATGTGGCGCGTATTATCCGCTGTGACGCGCGGGAAGAGCAACTCGTGGCGCAT 420
 Db 361 GCTATGTGGCGCGTATTATCCGCTGTGACGCGCGGGAAGAGCAACTCGTGGCGCAT 420
 QY 421 ACACTATTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGA 480
 Db 421 ACACTATTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGA 480
 QY 481 TGGCATCAGATGAAGAAAT 500
 Db 481 TGGCATCAGATGAAGAAAT 500

RESULT 2
 ACC44723
 ID ACC44723 standard; DNA; 2847 BP.
 AC ACC44723;
 XX
 XX
 XX 29-MAY-2003 (first entry)
 XX Vector pLIT38attB nucleotide sequence SEQ ID NO:120.
 XX
 XX Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
 KW att site; integrase; recombinase; ACes; gene therapy; transgenic animal;
 KW platform artificial chromosome expression system; gene; ds.
 XX
 XX Synthetic.
 XX
 XX WO200297059-A2.

XX PD 05-DEC-2002.
 XX PF 30-MAY-2002; 2002WO-US017452.
 XX PR 30-MAY-2001; 2001US-0294758P.
 XX PR 21-MAR-2002; 2002US-0366891P.
 XX (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
 XX Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
 XX Stewart S, Shellard J;
 XX WPI; 2003-140461/13.
 XX
 XX Novel eukaryotic chromosome comprising one or many att sites which
 XX permits site-directed integration in the presence of lambda-integrase,
 XX useful for site-specific recombination-directed integration of DNA of
 XX interest.
 XX
 XX Example 10; Page 257-258; 272pp; English.
 XX
 XX The present invention describes a eukaryotic chromosome (I) comprising
 XX one or several att sites, where an att site is heterologous to the
 XX chromosome, and permits site-directed integration in the presence of
 XX lambda-integrase. Also described: (1) a platform artificial chromosome
 XX expression system (ACes) (II) comprising several sites that participate
 XX in recombinase catalysed recombination; and (2) a method (M1) for
 XX introducing a heterologous nucleic acid into a platform artificial
 XX chromosome. (I) can be used in gene therapy. (M1) is useful for
 XX introducing a heterologous nucleic acid molecule into a platform
 XX artificial chromosome, preferably an ACes. (II) is useful for producing a
 XX transgenic animal (e.g. a fish, insect, reptile, amphibian, arachnid, or
 XX mammal) by introducing (II) by cell fusion, lipid-mediated transfection,
 XX by a carrier system, microinjection, microcell fusion, electroporation,
 XX microprojectile bombardment or direct DNA transfer into an embryonic
 XX cell, preferably a stem cell or an embryo. (II) comprises a heterologous
 XX nucleic acid that encodes a therapeutic product which is useful for
 XX making a library of ACes comprising random portions of a genome. ACC44612
 XX to ACC44732 and ABP96650 to ABP96657 represent sequences used in the
 XX exemplification of the present invention
 XX
 XX Sequence 2847 BP; 742 A; 685 C; 720 G; 700 T; 0 U; 0 Other;

Query Match 100.0%; Score 500; DB 7; Length 2847;
 Best Local Similarity 100.0%; Pred. No. 3.6e-124;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAACTACGTGAGTGGGACATTTTCGGGGAATGTGGCGGAACCCCTATTGTTATT 60
 Db 1 GTTAACTACGTGAGTGGGACATTTTCGGGGAATGTGGCGGAACCCCTATTGTTATT 60
 QY 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGACATTAACCTGATTAATCTTCA 120
 Db 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGACATTAACCTGATTAATCTTCA 120
 QY 121 ATAATATTGAAAAGGAAGATGATGATTCAACATTTCCGTCGCGCCCTATTCCCTT 180
 Db 121 ATAATATTGAAAAGGAAGATGATGATTCAACATTTCCGTCGCGCCCTATTCCCTT 180
 QY 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGA 240
 Db 181 TTTTGGCGCATTTTGGCTTCCTGTTTTCCTACCCAGAAACGCTGGTGAAGTAAAGA 240
 QY 241 TGTGTAAGATCAGTGGGTGACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA 300
 Db 241 TGTGTAAGATCAGTGGGTGACAGTGGGTACATCGAACTGGATCTCAACAGCGGTAA 300
 QY 301 GATCCTTGAGAGTTTTCGCCCGAAGACGTTCTCCATGATGAGACACTTTTAAAGTCT 360
 Db 301 GATCCTTGAGAGTTTTCGCCCGAAGACGTTCTCCATGATGAGACACTTTTAAAGTCT 360
 QY 361 GCTATGTGGCGCGTATTATCCGCTGTGACGCGCGGGAAGAGCAACTCGTGGCGCAT 420

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:24:05 ; Search time 434.846 Seconds
(without alignments)
4884.713 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500

Perfect score: 500
Sequence: 1 gttactactgcagtgga.....tggcatgacactagaagaat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	2814	7	ACC44722 Vector pL
2	500	100.0	2847	7	ACC44723 Vector pL
3	500	100.0	3419	3	AAAI4902 Nucleotid
4	500	100.0	3438	7	ACC44695 Plasmid p
5	500	100.0	3438	7	ABT16608 Artificia
6	500	100.0	3476	2	AAV33299 E.coli-Co
7	500	100.0	3476	6	ABK13447 Plasmid p
8	500	100.0	4223	7	ACC44724 Plasmid p
9	500	100.0	4223	7	ACC44714 Plasmid p
10	500	100.0	4768	4	AA05243 Plasmid p
11	500	100.0	4768	6	ABL42019 Nucleotid
12	500	100.0	4768	6	AB553351 Plasmid v
13	500	100.0	4768	6	AA517143 Gene targ
14	500	100.0	4768	6	AA028659 Plasmid p
15	500	100.0	5759	6	ABK49521 Positive
16	500	100.0	6148	6	ABK49520 Positive
17	500	100.0	6232	5	AAI16796 Chloropla
18	500	100.0	6477	5	AAI16797 Chloropla
19	500	100.0	6861	3	AA097521 Plasmid p
20	497	99.4	6355	4	AA05244 Plasmid v
21	497	99.4	6355	6	ABL42020 Nucleotid
22	497	99.4	6355	6	AB553352 Plasmid v
23	497	99.4	6355	6	AA517144 Gene targ

ALIGNMENTS

RESULT 1

ACC44722
ID ACC44722 standard; DNA; 2814 BP.
XX
AC ACC44722;
XX
DT 29-MAY-2003 (first entry)
XX
DE Vector pLITMUS 38 nucleotide sequence SEQ ID NO:119.
XX
XX Chromosome-based platform; artificial chromosome; eukaryotic chromosome;
att site; integrase; recombinase; ACes; gene therapy; transgenic animal;
XX
XX Platform artificial chromosome expression system; Gene; ds.
XX
XX Synthetic.
XX
XX WO200297059-A2.
XX
PD 05-DEC-2002.
XX
XX 30-MAY-2002; 2002WO-US017452.
XX
PR 30-MAY-2001; 2001US-0294758P.
PR 21-MAR-2002; 2002US-0366891P.
XX
XX (CHRO-) CHROMOS MOLECULAR SYSTEMS INC.
XX
XX Perkins E, Perez C, Lindenbaum M, Greene A, Leung J, Fleming E;
XX Stewart S, Shellard J;
XX WPI; 2003-140461/13.

24 497 99.4 6355 6 AAD28660 Plasmid p
25 497 99.4 6503 3 AAA12622 Nucleotid
26 496.8 99.4 3166 7 ACC44717 Base vect
27 496.8 99.4 4615 7 ACC44720 Plasmid p
28 496.8 99.4 6113 7 ACC44729 Plasmid p
29 496.8 99.4 7600 7 ACC44718 Plasmid p
30 496.8 99.4 7631 7 ACC44719 Plasmid p
31 496.8 99.4 8851 7 ACC44727 Plasmid p
32 496.8 99.4 9080 7 ACC44713 Plasmid p
33 496.8 99.4 10474 7 ACC44728 Plasmid p
34 495.4 99.1 8893 6 ABA05547 Maxadlan
35 495.2 99.0 3476 2 AAV48232 Vector pl
36 492.4 98.5 5264 7 ACC43141 Nucleotid
37 492.4 98.5 5408 7 ACC43140 Nucleotid
38 492.4 98.5 8590 2 AAT35168 Plasmid p
39 492 98.4 6714 2 AAQ79545 Plasmid p
40 492 98.4 6714 6 AAD42537 Plasmid p
41 492 98.4 6714 8 ACH00379 Plasmid p
42 492 98.4 8475 4 AAF76964 Genetic c
43 492 98.4 8492 4 AAF76963 Genetic c
44 492 98.4 8537 4 AAF76962 Genetic c
45 491.8 98.4 15065 3 AA236195 Nucleotid

XX Novel eukaryotic chromosome comprising one or many att sites which
permits site-directed integration in the presence of lambda-integrase,
XX useful for site-specific recombination-directed integration of DNA of
XX interest.
XX
XX Example 10; Page 256-257; 272pp; English.
XX
XX The present invention describes a eukaryotic chromosome (I) comprising
one or several att sites, where an att site is heterologous to the
XX chromosome, and permits site-directed integration in the presence of
XX lambda-integrase. Also described: (1) a platform artificial chromosome
XX expression system (Acce) (II) comprising several sites that participate
XX in recombinase catalysed recombination; and (2) a method (M1) for
XX introducing a heterologous nucleic acid into a platform artificial

Db 4221 ATAATATTGAAAAGGAGAGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTT 4162
QY 181 TTTTGGGGCATTTTGGCTTCTCTCTTTTGGCTCACCCAGAAACGCTGGTCAAAAGTAAAGA 240
Db 4161 TTTTGGGGCATTTTGGCTTCTCTCTTTTGGCTCACCCAGAAACGCTGGTCAAAAGTAAAGA 4102
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTACATCGAACTCGATCTCAACAGCGGTAA 300
Db 4101 TGCTGAAGATCAGTTGGGTGACGAGTGGGTACATCGAACTCGATCTCAACAGCGGTAA 4042
QY 301 GATCCTTGAGAGTTTTCGCCCGGAGAAAGTTTCCAAATGATGAGCACTTTTAAAGTTCT 360
Db 4041 GATCCTTGAGAGTTTTCGCCCGGAGAAAGTTTCCAAATGATGAGCACTTTTAAAGTTCT 3982
QY 361 GCTATCTGGCGCGTATTATCCCGTCTTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 420
Db 3981 GCTATCTGGCGCGTATTATCCCGTCTTGACCGCGGCAAGAGCAACTCGGTGCGCGCAT 3922
QY 421 ACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCAAGAAAGCATCTTACCGA 480
Db 3921 ACACTATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCAAGAAAGCATCTTACCGA 3862
QY 481 TGGCATGACAGTAAGAGAAT 500
Db 3861 TGGCATGACAGTAAGAGAAT 3842

Search completed: April 22, 2004, 11:46:26
Job time : 2798.02 secs

```
QY 121 ATAAATATTGAAAAGGAGAGATGATGATTTCAACATTTCCGTGCGCCCTTATTTCCCTT 180
Db 1406 ATAAATATTGAAAAGGAGAGATGATGATTTCAACATTTCCGTGCGCCCTTATTTCCCTT 1455
QY 181 TTTTGGGGCATTTTGGCTTCTGTTTTCCTCAACAGAAACGCTGGTGAAGTAAAGA 240
Db 1466 TTTTGGGGCATTTTGGCTTCTGTTTTCCTCAACAGAAACGCTGGTGAAGTAAAGA 1525
QY 241 TCTCAAGATCAGTTGGGTGACGAGTGGTTCATCGAATCGGATCTCAACAGCGTAA 300
Db 1526 TCTCAAGATCAGTTGGGTGACGAGTGGTTCATCGAATCGGATCTCAACAGCGGTA 1585
QY 301 GATCTTGAGATTTTCCGCCCGAAGACGTTCTCCAATGATGACATTTTAAAGTTCT 360
Db 1586 GATCTTGAGATTTTCCGCCCGAAGACGTTTCCCAATGATGACATTTTAAAGTTCT 1645
QY 361 GCTATGTGCGGGTATTATTCCTGTTGACGCGGGCAAGACGAACTCGGTGCGCGCAT 420
Db 1646 GCTATGTGCGGGTATTATTCCTGTTGACGCGGGCAAGACGAACTCGGTGCGCGCAT 1705
QY 421 ACATCTATTCTCAGATGACTTGGTTGAGTACTACCACTCAGAGAAAGCATCTTACGGA 480
Db 1706 ACATCTATTCTCAGATGACTTGGTTGAGTACTACCACTCAGAGAAAGCATCTTACGGA 1765
QY 481 TGGCATGACAGTAAGAGAT 500
Db 1766 TGGCATGACAGTAAGAGAT 1785

RESULT 14
LOCUS I63563 8590 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 70 from patent US 5663143.
ACCESSION I63563
VERSION I63563.1 GI:2481136
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 8590)
AUTHORS Ley,A.Charles., Ladner,R.Charles., Guterman,S.Kosow.,
TITLE Roberts,B.Lindsay., Markland,W. and Kent,R.Baribault.
JOURNAL Engineered human-derived kunitz domains that inhibit human
FEATURES
source
location/Qualifiers
1..8590
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 98.5%; Score 492.4; DB 6; Length 8590;
Best Local Similarity 99.8%; Pred. No. 2.2e-126;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 TACGTACAGTGGCATTTCGGGGAATGTGCGGGAACCCCTATTTCCTTTTCTA 66
Db 5931 TACGTACAGTGGCATTTCGGGGAATGTGCGGGAACCCCTATTTCCTTTTCTA 5990
QY 67 ATATCATTCATATGATTCCTCATGAGCAATAACCTGATTAATGCTCAATA 126
Db 5991 ATATCATTCATATGATTCCTCATGAGCAATAACCTGATTAATGCTCAATA 6050
QY 127 TTGAAAAGGAAGATGATGATTTCAACATTTCCGTGTCGCCCTTATTCCTTTTTCG 186
Db 6051 TTGAAAAGGAAGATGATGATTTCAACATTTCCGTGTCGCCCTTATTCCTTTTTCG 6110
QY 187 GCGATTTTCCCTTCTGTTTTCCTCAACAGAAACGCTGGTGAAGTAAAGATGCTGA 246
Db 6111 GCGATTTTCCCTTCTGTTTTCCTCAACAGAAACGCTGGTGAAGTAAAGATGCTGA 6170
QY 247 AGATCATGTTGGTGCACGAGTGGGTATACATCGAATCGATCTCAACAGCGGTAAGATCCT 306
```

```
Db 6171 AGATCATGTTGGTGCACGAGTGGGTATACATCGAATCGATCTCAACAGCGGTAAGATCCT 6230
QY 307 TGAGAGTTTTCGCCCGAAGACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCTATG 366
Db 6231 TGAGAGTTTTCGCCCGAAGACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATG 6290
QY 367 TGGCGGGTATTATCCCGTGTTCACCGCGGCAAGACAACTCGGTGCGCGCATACACTA 426
Db 6291 TGGCGGGTATTATCCCGTGTTCACCGCGGCAAGACAACTCGGTGCGCGCATACACTA 6350
QY 427 TTCTCAGATGACTTGGTTGAGTACTACCACTCAGAGAAAGCATCTTACGATGCGAT 486
Db 6351 TTCTCAGATGACTTGGTTGAGTACTACCACTCAGAGAAAGCATCTTACGATGCGAT 6410
QY 487 GACAGTAAGAGAT 500
Db 6411 GACAGTAAGAGAT 6424

RESULT 15
LOCUS XXU16722/c 4779 bp DNA circular SYN 24-MAY-1995
DEFINITION Cloning vector pLACT7, complete sequence.
ACCESSION U16722
VERSION U16722.1 GI:565104
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 4779)
AUTHORS Chong,S. and Garcia,G.A.
TITLE A versatile and general prokaryotic expression vector, pLACT7
JOURNAL Biotechniques 17 (4), 686-691 (1994)
MEDLINE 95134409
PUBMED 7833029
REFERENCE 2 (bases 1 to 4779)
AUTHORS Garcia,G.A.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1994) George A. Garcia, College of Pharmacy,
University of Michigan, 428 Church St., Ann Arbor, MI 48109-1065
FEATURES
source
location/Qualifiers
1..4779
/organism="Cloning vector pLACT7"
/mol_type="genomic DNA"
/db_xref="taxon:37153"
terminator 170..220
/notes="from T7"
misc_feature 221..289
/notes="multiple cloning site"
misc_feature 284
/notes="NdeI site"
RBS 295..303
promoter 312..328
/notes="from T7"
promoter 362..386
/notes="for lac gene"

ORIGIN
Query Match 98.4%; Score 492; DB 12; Length 4779;
Best Local Similarity 99.0%; Pred. No. 2.7e-126;
Matches 495; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTTAACTAGTCAGTGGCATTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 60
Db 4341 GTTAACTAGTCAGTGGCATTTCGGGGAATGTGCGGGAACCCCTATTGTTTATT 4282
QY 61 TTTCTAAATACATTCAATATGATCGCTCATGAGCAATAACCTGATTAATGCTTCA 120
Db 4281 TTTCTAAATACATTCAATATGATCGCTCATGAGCAATAACCTGATTAATGCTTCA 4222
QY 121 ATAATATTGAAAAGGAAGATGATGATTTCAACATTTCCGTGTCGCCCTTATTCCTT 180
```

```
QY 4 AACTACGTCAGTGGCACTTTTCGGGGAATGTCGGCGGAACCCCTATTGTTTATTTT 63
Db 1 AACTACGTCAGTGGCACTTTTCGGGGAATGTCGGCGGAACCCCTATTGTTTATTTT 60
QY 64 CTAATACATTCATATGATCGCTCATGAGACATTAACCTGATATAATGCTTCAATA 123
Db 61 CTAATACATTCATATGATCGCTCATGAGACATTAACCTGATATAATGCTTCAATA 120
QY 124 ATATTGAAAAGGAGATGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTT 183
Db 121 ATATTGAAAAGGAGATGATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTT 180
QY 184 TCGGCATTTTCCTTCTCTGTTTTCCTCACCAGAACGCTGCTGAAGTAAAGATGC 243
Db 181 TCGGCATTTTCCTTCTCTGTTTTCCTCACCAGAACGCTGCTGAAGTAAAGATGC 240
QY 244 TGAAGATCAGTTGGGTGACGAGTGGGTATCATCGAATCGGATCTCAACAGCGTAAGAT 303
Db 241 TGAAGATCAGTTGGGTGACGAGTGGGTATCATCGAATCGGATCTCAACAGCGTAAGAT 300
QY 304 CTTTACAGATTTTCGCCCGAGACGTTCTCCATGATGACGACTTTTAAAGTTCGCT 363
Db 301 CTTTACAGATTTTCGCCCGAGACGTTCTCCATGATGACGACTTTTAAAGTTCGCT 360
QY 364 ATGTGCGCGGTATTATCCGTTGTCAGCGCGGCAAGAGCAACTCGCTCGCCGATACA 423
Db 361 ATGTGCGCGGTATTATCCGTTGTCAGCGCGGCAAGAGCAACTCGCTCGCCGATACA 420
QY 424 CTATTTCTAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAGCATCTTACGGATGG 483
Db 421 CTATTTCTAGATGACTTGGTTGAGTACTCACCAGTCAAGAAAGCATCTTACGGATGG 480
QY 484 CATGACAGTAAGAGAT 500
Db 481 CATGACAGTAAGAGAT 497

RESULT 13
AY123047
LOCUS Erwinia amylovora plasmid pEA2.8 complete sequence.
DEFINITION Erwinia amylovora plasmid pEA2.8 complete sequence.
ACCESSION AY123047
VERSION AY123047.1 GI:26245316
KEYWORDS
SOURCE Erwinia amylovora
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Erwinia amylovora
REFERENCE 1 (bases 1 to 2825)
AUTHORS McChes, G.C., Schnabel, E.L., Maxson-Stein, K., Jones, B.,
Stromberg, V.K., Lacy, G.H. and Jones, A.L.
TITLE Relatedness of Chromosomal and Plasmid DNAs of Erwinia pyrifoliae
and Erwinia amylovora
JOURNAL Appl. Environ. Microbiol. 68 (12), 6182-6192 (2002)
PUBLISHED 12450843
REFERENCE 2 (bases 1 to 2825)
AUTHORS McChes, G.C., Schnabel, E.L., Maxson-Stein, K., Jones, B.,
Stromberg, V.K., Lacy, G.H. and Jones, A.L.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Plant Pathology, Michigan State University,
East Lansing, MI 48824, USA
FEATURES
source
1..2825
/organism="Erwinia amylovora"
/mol_type="genomic DNA"
/strain="IL-5"
/species_host="Rubus"
/db_xref="taxon:552"
/plasmid="pEA2.8"
/country="USA"
127..132
/evidence=not_experimental

-10_signal
misc_feature
gene
misc_RNA
-10_signal
-35_signal
CDS
rep_origin
CDS
CDS
CDS
ORIGIN
Query Match 99.4%; Score 496.8; DB 1; Length 2825;
Best Local Similarity 99.6%; Pred. No. 1.2e-127;
Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GTTAACTACGTCAGTGGCACTTTTCGGGGAATGTCGGCGGAACCCCTATTGTTTATTT 60
Db 1286 GTTAACTACGTCAGTGGCACTTTTCGGGGAATGTCGGCGGAACCCCTATTGTTTATTT 1345
QY 61 TTCTTAATACATTCATATGATGATCGCTCATGAGACATTAACCTGATATAATGCTTCA 120
Db 1346 TTCTTAATACATTCATATGATGATCGCTCATGAGACATTAACCTGATATAATGCTTCA 1405

-35_signal
```


Query Match 99.4%; Score 497; DB 6; Length 6355;
Best Local Similarity 100.0%; Pred. No. 1.1e-127; Indels 0; Gaps 0;
Matches 497; Conservative 0; Mismatches 0;

QY 4 AACTACGTCAGGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTATTTT 63
DB 1591 AACTACGTCAGGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTATTTT 1650

QY 64 CTAAATACATTCAAATATGATCCGCTCATGAGACAATAACCCGTGATAAATGCTTCAATA 123
DB 1651 CTAAATACATTCAAATATGATCCGCTCATGAGACAATAAACCCTGATAAATGCTTCAATA 1710

QY 124 ATATTGAAAAGAGAGATGATGAGTATTCACATTTCCGTCGCGCTTATTCCTTTT 183
DB 1711 ATATTGAAAAGAGAGATGATGAGTATTCACATTTCCGTCGCGCTTATTCCTTTT 1770

QY 184 TCGGCAATTTCCCTTCCTGTTTTCCTTCACCCAGAAACGCTGGTGAAGTAAAGATGC 243
DB 1771 TCGGCAATTTCCCTTCCTGTTTTCCTTCACCCAGAAACGCTGGTGAAGTAAAGATGC 1830

QY 244 TGAAGATCAGTTGGGTGACGAGTGGTTTACATCGAATGGAATCTCAACAGCGGTAAGAT 303
DB 1831 TGAAGATCAGTTGGGTGACGAGTGGTTTACATCGAATGGAATCTCAACAGCGGTAAGAT 1890

QY 304 CTTTGAGATTTTCGCCCGAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTCTGCT 363
DB 1891 CTTTGAGATTTTCGCCCGAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTCTGCT 1950

QY 364 ATGTGGCGGCTATTATCCGCTGTTGACCGCGGGAAGAGCAACTCGGTCGCGCATACA 423
DB 1951 ATGTGGCGGCTATTATCCGCTGTTGACCGCGGGAAGAGCAACTCGGTCGCGCATACA 2010

QY 424 CTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCACTTACGATGG 483
DB 2011 CTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCACTTACGATGG 2070

QY 484 CATGACAGTAAGAGAT 500
DB 2071 CATGACAGTAAGAGAT 2087

RESULT 11
AX352705
LOCUS AX352705 6355 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 2 from Patent WO0204621.
ACCESSION AX352705
VERSION AX352705.1 GI:18617826
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Klein, R.D. and Brennan, T.J.
TITLE Methods of creating constructs useful for introducing sequences in to embryonic stem cells
JOURNAL Patent: WO 0204621-A 2 17-JAN-2002;
Deltagen, Inc. (US)
FEATURES
source
1. .6355
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Plasmid vector"

ORIGIN
Query Match 99.4%; Score 497; DB 6; Length 6355;
Best Local Similarity 100.0%; Pred. No. 1.1e-127; Indels 0; Gaps 0;
Matches 497; Conservative 0; Mismatches 0;

QY 4 AACTACGTCAGGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTATTTT 63
DB 1591 AACTACGTCAGGTGGCACTTTTCGGGAAATGTGCGGAAACCCCTATTGTTATTTT 1650

QY 64 CTAAATACATTCAAATATGATCCGCTCATGAGACAATAACCCGTGATAAATGCTTCAATA 123
DB 1651 CTAAATACATTCAAATATGATCCGCTCATGAGACAATAAACCCTGATAAATGCTTCAATA 1710

QY 124 ATATTGAAAAGAGAGATGATGAGTATTCACATTTCCGTCGCGCTTATTCCTTTT 183
DB 1711 ATATTGAAAAGAGAGATGATGAGTATTCACATTTCCGTCGCGCTTATTCCTTTT 1770

QY 184 TCGGCAATTTCCCTTCCTGTTTTCCTTCACCCAGAAACGCTGGTGAAGTAAAGATGC 243
DB 1771 TCGGCAATTTCCCTTCCTGTTTTCCTTCACCCAGAAACGCTGGTGAAGTAAAGATGC 1830

QY 244 TGAAGATCAGTTGGGTGACGAGTGGTTTACATCGAATGGAATCTCAACAGCGGTAAGAT 303
DB 1831 TGAAGATCAGTTGGGTGACGAGTGGTTTACATCGAATGGAATCTCAACAGCGGTAAGAT 1890

QY 304 CTTTGAGATTTTCGCCCGAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTCTGCT 363
DB 1891 CTTTGAGATTTTCGCCCGAAGACGTTCTCCAAATGATGAGCACTTTTAAAGTCTGCT 1950

QY 364 ATGTGGCGGCTATTATCCGCTGTTGACCGCGGGAAGAGCAACTCGGTCGCGCATACA 423
DB 1951 ATGTGGCGGCTATTATCCGCTGTTGACCGCGGGAAGAGCAACTCGGTCGCGCATACA 2010

QY 424 CTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCACTTACGATGG 483
DB 2011 CTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCACTTACGATGG 2070

QY 484 CATGACAGTAAGAGAT 500
DB 2071 CATGACAGTAAGAGAT 2087

RESULT 12
BD249689
LOCUS BD249689 6509 bp DNA linear PAT 17-JUL-2003
DEFINITION Intein mediated peptide ligation.
ACCESSION BD249689
VERSION BD249689.1 GI:33059459
KEYWORDS JP 2002526387-A/5.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 6509)
AUTHORS Xu, M.Q. and Evans, T.C.
TITLE Intein mediated peptide ligation
JOURNAL Patent: JP 2002526387-A 5 20-AUG-2002;
NEW ENGLAND BIOLABS INC
COMMENT OS Artificial Sequence
PN JP 2002526387-A/5
PD 20-AUG-2002
PF 30-SEP-1999 JP 2000572329
PR 30-SEP-1998 US 60/102413
PI MING QUN XU, THOMAS C EVANS
PC C07K19/00,C07K1/04,C07K1/07,C12N15/09,C12N15/09,C12N15/00, PC C12N15/00
CC Description of Artificial Sequence: pTXB1 plasmid sequence CC containing the
CC modified intein from the gyrA gene of Mycobacterium xenopi FH
Key Location/Qualifiers
FT source
1. .6509
/organism="Artificial Sequence".
Location/Qualifiers
1. .6509
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Query Match 99.4%; Score 497; DB 6; Length 6509;
Best Local Similarity 100.0%; Pred. No. 1.1e-127; Indels 0; Gaps 0;
Matches 497; Conservative 0; Mismatches 0;

TITLE Construction and use of low-copy number T7 expression vectors for purification of problem proteins: affinity purification of Mycobacterium tuberculosis RmD and Pseudomonas aeruginosa Lasi and Rhl proteins, and functional analysis of RhlI

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 6278)

AUTHORS Hoang, T.T. and Schweizer, H.P.

TITLE Direct Submission

JOURNAL Submitted (30-APR-1999) Microbiology, Colorado State University, Fort Collins, CO 80523, USA

FEATURES

Source

1. 6278

/organism="Expression vector pNam"

/mol_type="genomic DNA"

/db_xref="taxon:101373"

complement(144..1004)

/gene="bla"

complement(144..1004)

/gene="bla"

/codon_start=1

/transl_table=11

/product="beta-lactamase"

/protein_id="AAD49603.1"

/db_xref="GI:5733490"

/translation="MSIQHFRVALIPFFAFCPLPVFAHPETLVKVKDAEDQLGARVGY IELDLNSGKILSPKPEERSPMWSTFKVLLCGAVLSRVDAGQEOQLGRHHYSONDLVE YSPVTEKHLTDGKTVRELCSAAITWSDNTAANLLLTIGGPKELTAPLHWGNDVTRL DREPELNAIPENDSRDTPMVAMATLRLKLTGELLTLASRQQLIDWQADKVAQPL LRSAUPAGWFIADKSGAGERSGIIALGPDPKPSRIWVITGSGQATMDERARQIA EIGASLIQHW"

complement(1292..>1450)

/gene="cbd"

complement(1292..>1450)

/gene="cbd"

/codon_start=1

/transl_table=11

/product="chitin binding domain"

/protein_id="AAD49604.1"

/db_xref="GI:5733491"

/translation="TTNPGVSAWQVNTAYTAGQLVTYNGKTYKICLPHTSLAGWEPNS VPALWQLQ"

complement(<1484..>2845)

/gene="int"

complement(<1484..>2845)

/gene="int"

/codon_start=1

/transl_table=11

/product="intein"

/protein_id="AAD49605.1"

/db_xref="GI:5733492"

/translations="CPAKTNVLMADGSIETCENIEVGNKVMGKDGPREVTKLPGR ETWYSVQKSHRAKSDSREVPKLLKFTCNATHELVTSPRSVRLSRITKGVYF EVITFEMGQKAPDGRIVELVPEKSYFISEGPERANELVESYKRNAGKAYFEVIE ARDLISLGHVRKATQYAPILYENDHFDYQKSKPHLTIEGPKVAYLLGLWIG GLSDRATFSVSDRTSLMERVYAEKLNLCAYKDRKEPQVAKTVNLYSKVRNGI RNNLTENPLWDAIVGLGKDGKVNIPSLDNLGTRETFLAGLIDSDGYVTDHG IKATIKTHTSVSDGLVSLARSLGLVSNAREPAKVMNVTKHKSYSATVYMGSGDVL NLSKAGSKKFPAPAPARECRGCFYELQELKEDDDYIGITLSDSDSHQFLQSGVQ VVHA"

2846..2899

/note="multiple cloning site"

2969..2984

/note="T7 promoter"

3375..4457

/gene="lacI"

3375..4457

/gene="lacI"

/codon_start=1

/transl_table=11

/product="lac repressor"

/protein_id="AAD49606.1"

/db_xref="GI:5733493"

/translation="MKPVTLLVDVNEYAGVSYQTVSRVNVQASHVSAKTRKVEANMAE

misc_feature

2846..2899

/note="multiple cloning site"

2969..2984

/note="T7 promoter"

3375..4457

/gene="lacI"

3375..4457

/gene="lacI"

/codon_start=1

/transl_table=11

/product="lac repressor"

/protein_id="AAD49606.1"

/db_xref="GI:5733493"

/translation="MKPVTLLVDVNEYAGVSYQTVSRVNVQASHVSAKTRKVEANMAE

promoter

2969..2984

/note="T7 promoter"

gene

3375..4457

/gene="lacI"

CDS

3375..4457

/gene="lacI"

misc_feature

4829..6277

/note="pSC101 replicon"

ORIGIN

Query Match 99.7%; Score 498.4; DB 12; Length 6278;

Best Local Similarity 99.8%; Pred. No. 4.5e-128; Indels 0; Gaps 0;

Matches 499; Conservative 0; Mismatches 1;

QY 1 GTTAACTAGCTCAGGTGGCCTTTTCGGGGAATATGTCGGCGAACCCTTATTGTTTAT 60

Db 1146 GATAACTAGCTCAGGTGGCCTTTTCGGGGAATATGTCGGCGAACCCTTATTGTTTAT 1087

QY 61 TTCTTAAATACATTCAAATATGATATCGCTCATGAGCAATACCTGTAATATGCTTCA 120

Db 1086 TTCTTAAATACATTCAAATATGATATCGCTCATGAGCAATACCTGTAATATGCTTCA 1027

QY 121 ATATATTTGAAAAGCAAGATGATGATTAACAATTTCCGTGTCGCCCTTATTCCTT 180

Db 1026 ATAATATTTGAAAAGCAAGATGATGATTAACAATTTCCGTGTCGCCCTTATTCCTT 967

QY 181 TTTTGGGCAATTTTGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240

Db 966 TTTTGGGCAATTTTGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 907

QY 241 TGCTGAAGATCAGTTGGTGCACAGTGGGTACATCGAATCGAATCGAATCGAATCGAAT 300

Db 906 TGCTGAAGATCAGTTGGTGCACAGTGGGTACATCGAATCGAATCGAATCGAATCGAAT 847

QY 301 GATCCTTTGAGAGTTTTCGCCCGCAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 360

Db 846 GATCCTTTGAGAGTTTTCGCCCGCAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCT 787

QY 361 GCTATGTGGCGCGTATTATTCCTGTTGACGCGCGGCAAGCACTCGTCCGCGCAT 420

Db 786 GCTATGTGGCGCGTATTATTCCTGTTGACGCGCGGCAAGCACTCGTCCGCGCAT 727

QY 421 ACATATTCTCAGATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCACTCTTACGGA 480

Db 726 ACATATTCTCAGATGACTTGGTGTAGTACTCACCAGTACACAGAAAGCACTCTTACGGA 667

QY 481 TGGCATGACAGTAAGAGAAAT 500

Db 666 TGGCATGACAGTAAGAGAAAT 647

RESULT 10

AX299822

LOCUS AX299822 6355 bp DNA linear PAT 26-NOV-2001

DEFINITION Sequence 2 from Patent WO0167855.

ACCESSION AX299822

VERSION AX299822.1 GI:17129313

KEYWORDS synthetic construct

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE 1

AUTHORS Allen, K.D., Guenther, C. and Phillips, R.

TITLE Transgenic mice containing targeted gene disruptions

JOURNAL Patent: WO 0167855-A 2 20-SEP-2001;

DELTAGEN, Inc. (US)

FEATURES

source

1..6355

/organism="synthetic construct"

/mol_type="unassigned DNA"

/db_xref="taxon:32630"

/note="Phage Vector"

ORIGIN

```
QY 361 GCTATGTCGGCGGATTTATCCCGTGTGACGCGCGGCAAGAGCAATCCGTCGCCGCAT 420
Db 361 GCTATGTCGGCGGATTTATCCCGTGTGACGCGCGGCAAGAGCAATCCGTCGCCGCAT 420
QY 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCACTACAGAAAAGCATCTTACGGA 480
Db 421 ACATATTCTCAGATGACTTGGTTGAGTACTACCACTACAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAAGAT 500
Db 481 TGGCATGACAGTAAGAAGAT 500

RESULT 8
AF397196/c
LOCUS AF397196 6293 bp DNA circular SYN 21-AUG-2001
DEFINITION Retrofittng vector pRetroBS, complete sequence.
ACCESSION AF397196
VERSION AF397196.1 GI:15216973
KEYWORDS
SOURCE
ORGANISM
RETIROFITTING VECTOR pRetroBS
ARTIFICIAL SEQUENCES; VECTORS.
REFERENCE
1 (bases 1 to 6293)
AUTHORS Wang, Z., Engler, P., Longacre, A. and Storb, U.
TITLE An efficient method for high-fidelity BAC/PAC retrofittng with a
selectable marker for mammalian cell transfection
JOURNAL Genome Res. 11 (1), 137-142 (2001)
MEDLINE 21068695
PUBMED 11156622
REFERENCE
2 (bases 1 to 6293)
AUTHORS Wang, Z., Engler, P., Longacre, A. and Storb, U.
TITLE Direct Submision
JOURNAL Submitted (05-JUN-2001) Mol. Genet. Cell Biol., University of
Chicago, 920 E. 58th St., Chicago, IL 60637, USA
FEATURES
source
1..6293
/organism="Retrofittng vector pRetroBS"
/mol_type="genomic DNA"
/db_xref="taxon:168860"
/notes="for in vivo retrofittng of BAC/PAC clones"
561..2363
/gene="cre"
561..2363
/gene="cre"
/transl_table=1
/product="GST-loxp-cre recombinase fusion protein"
/protein_id="AAK92454.1"
/db_xref="GI:15216974"

misc_feature
1332..2360
/gene="cre"
/notes="Region: cre recombinase"
complement(3064..3867)
/notes="for mammalian cell selection"
/codon_start=1
/transl_table=1
/product="PGK-neo"
/protein_id="AAK92455.1"
/db_xref="GI:15216975"
QGRPVLVFKTDLGALNELQDEARLSWLTGVPFAAALVDVTEAGRWLLLGEVPG
```

```
QDLLSHLPAEKVKSIMADAMERLHTLDPATCFPDHQAKHRIERARTRMEAGLVDDDD
LDEHQGLAPAEFLFARLKPMDGDLVVTTHGDACLPNIMVENGFSFGFIDCRLGVA
DRYQDIALATRDIAEELGGEWADRELVLGYAAPDSQRIAFVRLLEDFE"
4486..4905
/notes="R6K conditional replication origin"
complement(5032..5892)
/notes="ampicillin-resistance protein"
/codon_start=1
/transl_table=1
/product="beta-lactamase"
/protein_id="AAK92456.1"
/db_xref="GI:15216976"
/translation="MSIQHFRVALIPFFAECFLPVFAHPETLVKVKDAEDQLGARVGY
IETLNSKRILESFRRPEERSMMTSFKVLLCGNVLSRDVAGQGLGRHIIHQSDNLVE
YSEPTVKHLTDGMTVELCSAATMSDNTLANLLTTIGGPKELTAFVHNMGDHVTSL
DRWPELNEAIENDESDTTTPVMAMATLTKLLTGLLTLASQQIDIMWEADKVGAPL
LRSALPAGWFIADKSGAGSRGIIAALPGDKPSRIIVITVTSQATMDERNRQIA
EIGASLIKHW"

ORIGIN
Query Match 100.0%; Score 500; DB 12; Length 6293;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACTAGCTCAGGTGGCACTTTTCGGGCAATATGCGCGAACCCTTATTGTTTATT 60
Db 6034 GTTAACTAGCTCAGGTGGCACTTTTCGGGCAATATGCGCGAACCCTTATTGTTTATT 5975
QY 61 TTCTAAATACATTCAAATATGATCCGCTGATGAGCAATAACCCGTGATAAATGCTTCA 120
Db 5974 TTCTAAATACATTCAAATATGATCCGCTGATGAGCAATAACCCGTGATAAATGCTTCA 5915
QY 121 ATAATATTGAAAAGGAGAGTATGATGATTTCCGCTGCGCCCTTATTCCTT 180
Db 5914 ATAATATTGAAAAGGAGAGTATGATGATTTCCGCTGCGCCCTTATTCCTT 5855
QY 181 TTTTGGCGCATTTTGCCTTCCTGTTTCTCACCAGAAAACGCTGGTGAAGTAAAGA 240
Db 5854 TTTTGGCGCATTTTGCCTTCCTGTTTCTCACCAGAAAACGCTGGTGAAGTAAAGA 5795
QY 241 TGGTGAAGATCAGTTGGGTGCACGAGTGGTGTACATTCGAACCTGGATCTCAACAGCGGTA 300
Db 5794 TGGTGAAGATCAGTTGGGTGCACGAGTGGTGTACATTCGAACCTGGATCTCAACAGCGGTA 5735
QY 301 GATCCTTGAGAGTTTTCGCCCGCAAGAACGTTCTCCATGATGAGCACTTTTAAAGTTCT 360
Db 5734 GATCCTTGAGAGTTTTCGCCCGCAAGAACGTTCTCCATGATGAGCACTTTTAAAGTTCT 5675
QY 361 GCTATGTCGGCGGATTTATCCCGTGTTCACCGCGGCAAGAGCACTCGGTGCGCGCAT 420
Db 5674 GCTATGTCGGCGGATTTATCCCGTGTTCACCGCGGCAAGAGCACTCGGTGCGCGCAT 5615
QY 421 ACATATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 480
Db 5614 ACATATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCAAGAAAAGCATCTTACGGA 5555
QY 481 TGGCATGACAGTAAGAAGAT 500
Db 5554 TGGCATGACAGTAAGAAGAT 5535

RESULT 9
AF147463/c
LOCUS AF147463 6278 bp DNA circular SYN 16-AUG-1999
DEFINITION T7 Expression vector pNam, complete sequence.
ACCESSION AF147463
VERSION AF147463.1 GI:5733489
KEYWORDS
SOURCE
ORGANISM
RETIROFITTING VECTOR pNam
ARTIFICIAL SEQUENCES; VECTORS.
REFERENCE
1 (bases 1 to 6278)
AUTHORS Hoang, T.T., Stern, R.J., McNeil, M.R. and Schweizer, H.P.
```

QY 121 ATAAATATTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 180
DB 121 ATAAATATTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 180
QY 181 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 240
DB 181 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 240
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 300
DB 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 300
QY 301 GATCCTTTCAGAGTCTTCCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 360
DB 301 GATCCTTTCAGAGTCTTCCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 360
QY 361 GCTATGTGGCGGGTATTATCCGTTGTCAGCGCGGCAAGAGCAATCTCGTCCCGCAT 420
DB 361 GCTATGTGGCGGGTATTATCCGTTGTCAGCGCGGCAAGAGCAATCTCGTCCCGCAT 420
QY 421 ACACATATTCTCAGATGACTTGGTTGAGTACTCACCACTCAGAAAAGCATCTTACGGA 480
DB 421 ACACATATTCTCAGATGACTTGGTTGAGTACTCACCACTCAGAAAAGCATCTTACGGA 480
QY 481 TGGCATGACAGTAAGAGAT 500
DB 481 TGGCATGACAGTAAGAGAT 500

RESULT 6
AX468471
LOCUS AX468471 5100 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 14 from Patent WO222834.
ACCESSION AX468471
VERSION AX468471.1 GI:21901307
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 Siebel, C. and Brennan, T.J.
AUTHORS Methods of producing cells and animals comprising targeted gene
TITLE modifications
JOURNAL Patent: WO 0222834-A 14 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
source
1. .5100
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Construct Sequence"

ORIGIN
Query Match 100.0%; Score 500; DB 6; Length 5100;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACTACGTGAGTGGCACTTTTCGGGGAATGTGCGGAAACCCCTATTGTTTATT 60
DB 3296 GTTAACTACGTGAGTGGCACTTTTCGGGGAATGTGCGGAAACCCCTATTGTTTATT 3355
QY 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGCAATTAACCCGTGATAATGCTTCA 120
DB 3356 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGCAATTAACCCGTGATAATGCTTCA 3415
QY 121 ATAAATTTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 180
DB 3416 ATAAATTTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 3475
QY 181 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 240
DB 3476 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 3535

QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 300
DB 3536 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 3595
QY 301 GATCCTTTCAGAGTCTTTCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 360
DB 3596 GATCCTTTCAGAGTCTTTCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 3655
QY 361 GCTATGTGGCGGGTATTATCCGTTGTCAGCGCGGCAAGAGCAATCTCGTCCCGCAT 420
DB 3656 GCTATGTGGCGGGTATTATCCGTTGTCAGCGCGGCAAGAGCAATCTCGTCCCGCAT 3715
QY 421 ACACATATTCTCAGATGACTTGGTTGAGTACTCACCACTCAGAAAAGCATCTTACGGA 480
DB 3716 ACACATATTCTCAGATGACTTGGTTGAGTACTCACCACTCAGAAAAGCATCTTACGGA 3775
QY 481 TGGCATGACAGTAAGAGAT 500
DB 3776 TGGCATGACAGTAAGAGAT 3795

RESULT 7
AX468470
LOCUS AX468470 6148 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 13 from Patent WO222834.
ACCESSION AX468470
VERSION AX468470.1 GI:21901306
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 Siebel, C. and Brennan, T.J.
AUTHORS Methods of producing cells and animals comprising targeted gene
TITLE modifications
JOURNAL Patent: WO 0222834-A 13 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
source
1. .6148
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Construct Sequence"

ORIGIN
Query Match 100.0%; Score 500; DB 6; Length 6148;
Best Local Similarity 100.0%; Pred. No. 1.6e-128;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTAACTACGTGAGTGGCACTTTTCGGGGAATGTGCGGAAACCCCTATTGTTTATT 60
DB 1 GTTAACTACGTGAGTGGCACTTTTCGGGGAATGTGCGGAAACCCCTATTGTTTATT 60
QY 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGCAATTAACCCGTGATAATGCTTCA 120
DB 61 TTTCTAAATACATTCAAATATGATATCCGCTCATGAGCAATTAACCCGTGATAATGCTTCA 120
QY 121 ATAAATTTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 180
DB 121 ATAAATTTGAAAAGAGAGATGATGATTTCAACATTTCCGTCGCCCTTATTCCTT 180
QY 181 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 240
DB 181 TTTTCGGGCATTTTCCCTTCCCTGTTTTCGTCACCCAGAAACGCTGCTGAAAGTAAAGA 240
QY 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 300
DB 241 TGCTGAAGATCAGTTGGGTGACGAGTGGGTATCATGCAATGATCTCAACAGCGGTAA 300
QY 301 GATCCTTTCAGAGTCTTTCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 360
DB 301 GATCCTTTCAGAGTCTTTCGCCCCGAGAAACGCTTCCCAATGATGAGCACTTTTAAAGTTCT 360


```
Query Match 100.0%; Score 500; DB 6; Length 3476;
Best Local Similarity 100.0%; Pred. No. 1.5e-128;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:25:20 ; Search time 2795.02 Seconds
(without alignments)
7753.612 Million cell updates/sec

Title: US-09-954-483B-13_COPY_1_500
Perfect score: 500
Sequence: 1 gtaactactcagtggtgga.....tggcatgacagtagagaat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_man:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	500	100.0	3476	6	AR409373	AR409373 Sequence
C 2	500	100.0	3476	6	AR306561	AR306561 Sequence
C 3	500	100.0	3476	6	BD105910	BD105910 Immunotox
C 4	500	100.0	4768	6	AX299821	AX299821 Sequence
C 5	500	100.0	4768	6	AX352704	AX352704 Sequence
C 6	500	100.0	5100	6	AX468471	AX468471 Sequence
C 7	500	100.0	6148	6	AX468470	AX468470 Sequence
C 8	500	100.0	6293	12	AF397196	AF397196 Retrofitt
C 9	498.4	99.7	6278	12	AF147463	AF147463 T7 Expre
C 10	497	99.4	6355	6	AX299822	AX299822 Sequence
C 11	497	99.4	6355	6	AX352705	AX352705 Sequence
C 12	497	99.4	6509	6	BD249689	BD249689 Intein me
C 13	496.8	99.4	2825	1	AY123047	AY123047 Erwinia a
C 14	492.4	98.5	8590	6	I63563	I63563 Sequence 70
C 15	492	98.4	4779	12	XXU16722	XXU16722 Cloning vec
C 16	492	98.4	6714	6	E36448	E36448 DNA polymer
C 17	492	98.4	6714	6	I13349	I13349 Sequence 5
C 18	491.8	98.4	4109	12	SYNXYCTAP	M8536 Expression
C 19	491.8	98.4	4477	12	SYNXYVECHD	M8536 Expression
C 20	491.4	98.3	1054	1	AF104442	AF104442 Escherich
C 21	491.4	98.3	1075	1	PATN3PN1A	X54604 Pseudomonas
C 22	491.4	98.3	1086	6	AX212292	AX212292 Sequence
C 23	491.4	98.3	1106	6	I02541	I02541 Sequence 1
C 24	491.4	98.3	1652	6	I01971	I01971 Sequence 2
C 25	491.4	98.3	1905	6	AR027070	AR027070 Sequence
C 26	491.4	98.3	1905	6	I86203	I86203 Sequence 9
C 27	491.4	98.3	2048	12	SYNPOLY1A	M18125 Plasmid pPo
C 28	491.4	98.3	2061	12	SYNPOLY2A	M18128 Plasmid pPo
C 29	491.4	98.3	2117	12	SYNPOLY3I	M18131 Plasmid pPo
C 30	491.4	98.3	2320	6	AR371194	AR371194 Sequence
C 31	491.4	98.3	2320	6	BD08862	BD08862 Hepatitis
C 32	491.4	98.3	2369	12	AY230150	AY230150 Expression
C 33	491.4	98.3	2783	12	CVU47670	CVU47670 Cloning vec
C 34	491.4	98.3	2846	12	SYNOMPA	M59367 Cloning vec
C 35	491.4	98.3	2860	12	P8BC2	X68258 Bicistronic
C 36	491.4	98.3	2870	6	AX107930	AX107930 Sequence
C 37	491.4	98.3	3070	6	BD182167	BD182167 Cloning v
C 38	491.4	98.3	3097	6	AX107931	AX107931 Sequence
C 39	491.4	98.3	3122	6	AR161211	AR161211 Sequence
C 40	491.4	98.3	3122	6	AR370166	AR370166 Sequence
C 41	491.4	98.3	3122	6	BD096800	BD096800 Transgeni
C 42	491.4	98.3	3123	12	PMBC2TD	X77750 Plasmid pMB
C 43	491.4	98.3	3159	6	AX082146	AX082146 Sequence
C 44	491.4	98.3	3159	6	AX088825	AX088825 Sequence
C 45	491.4	98.3	3159	6	AX365120	AX365120 Sequence

ALIGNMENTS

RESULT 1
AR409373/c
LOCUS AR409373
DEFINITION Sequence 1 from patent US 6632928.
ACCESSION AR409373
VERSION AR409373.1 GI:40160295
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3476)
AUTHORS Neville,D.M. and Thompson,J.T.
TITLE Immunotoxins and methods of inducing immune tolerance
JOURNAL Patent: US 6632928-A 1 14-OCT-2003;
FEATURES Location/Qualifiers

3476 bp DNA linear PAT 18-DEC-2003

[illegible]

clones per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the battery of 200 other oligonucleotides are grouped into clusters. Clusters of 200 other oligonucleotides are grouped into clusters. Clusters of 200 other oligonucleotides are grouped into clusters.

library)"

/note="Organ: blood; Vector: pTriplex2; Site:1: Sfil;
Site 2: Sfil; A complementary DNA (cDNA) library from
human erythroid precursor cells was constructed using
SMART PCR (polymerase chain reaction) cDNA Library
Construction Kit (Clontech, Palo Alto, CA) according to
the manufacturer's directions, but with slight
modifications. Briefly, reverse transcription was
performed in the presence of 1 umol/L peptide nucleic acid
(PNA) oligos
(N-terminal)-biotin-GTC-CAC-CGG-AAG-CTT-G-(C-terminal) and
(N-terminal)-biotin-C(T/C)T-GAA-GTT-CTC-AGG-A-(C-terminal)
. Synthesized cDNA was digested with Sfil and
size-selected on a 1% agarose gel (>800bp). Large-scale
sequencing of the library was performed by the NIH
Intramural Sequencing Center (NISC;
Http://www.nisc.nih.gov/).

ORIGIN

Query Match 44.7%; Score 28.6; DB 13; Length 654;
Best Local Similarity 72.5%; Pred. No. 7e+02; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14;

QY 1 GAATTCCTGCCAGACCATGCCAAAAAGAGAGAAAGTTCATGAACCA 51
|||||
Db 565 GAATTAACTGCAACAGATGACAGGGAAGAAAGATAGGCCATGAACA 515
|||||

RESULT 4

B1512790 549 bp mRNA linear EST 08-APR-2002
LOCUS B160010A20B08.5 Bee Brain Normalized Library, B16 Apis mellifera
DEFINITION B160010A20B08 5', mRNA sequence.

ACCESSION B1512790.1 GI:15363164
VERSION B1512790.1
KEYWORDS EST.
SOURCE Apis mellifera (honeybee)

ORGANISM

Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apoidea; Apoidea;
Apidae; Apis.

REFERENCE

1 (bases 1 to 549)
Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.,
Annotated expressed sequence tags and cDNA microarrays for studies
of brain and behavior in the honey bee
Genome Res. 12 (4), 555-566 (2002)

JOURNAL

MEDLINE 21929762
PUBMED 11932240

COMMENT

Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499

Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers

FORWARD: TAATACGACTCATATAGG

BACKWARD: ATTACCTCTACTAAG

Plate: B160010A20 row: B column: 08

Seq primer: AGCGATACATTTCACACGGA

High quality sequence stop: 549.

Location/Qualifiers

FEATURES

1..549

/organism="Apis mellifera"

/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly

A.m. ligustica"

/db_xref="taxon:7460"

/clone="B160010A20B08"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/clone_lib="Bee Brain Normalized Library, B16"

/note="Organ: brain; Vector: pTriplex2; Site 1: EcoRI;

Site 2: NotI; The B16 library was constructed by the

Soares laboratory and it was constructed and normalized

as described by Bonaldo, M.F., Lennon, G. and Soares,

M.B. (1996), genome Research 6(9): 791-806. RNA was

prepared from dissected brains of adult worker bees of

various ages and various behavioral groups."

ORIGIN

Query Match 44.1%; Score 28.2; DB 12; Length 549;

Best Local Similarity 68.4%; Pred. No. 9.2e+02; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 16;

QY 3 ATTCACTGCCAGACCATGCCAAAAAGAGAGAAAGTTCATGAACCAAGTTCACGTT 59

153 AATTACTTACTAAGACATTCCTCAAGAAAAACAGTAGAATCATCAACCAATTCAGTT 209

RESULT 5

FR0013439 288 bp DNA linear GSS 18-SEP-1997
LOCUS F.rubripes GSS sequence, clone 122014Bg11, genomic survey sequence.

DEFINITION

AL004689.1 GI:2450259

ACCESSION GSS; genome survey sequence.

KEYWORDS Takifugu rubripes (Fugu rubripes)

SOURCE

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodontidae; Takifugu.

1 (bases 1 to 288)

Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umranaia,Y.,

Williams,G. and Brenner,S.

Direct Submission

TITLE

Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.mrc.ac.uk

COMMENT

Vector: pBluescript II KS

V type: Phagemid

PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES

Location/Qualifiers

1..288

/organism="Takifugu rubripes"

/mol_type="genomic DNA"

/db_xref="taxon:31033"

/clone="122014Bg11"

/clone_lib="cosmid 122014"

ORIGIN

Query Match 43.8%; Score 28; DB 29; Length 288;

Best Local Similarity 66.7%; Pred. No. 1.2e+03; Indels 0; Gaps 0;

Matches 40; Conservative 0; Mismatches 20;

QY 4 TTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGTTCATGAACCAAGTTCATAC 63

39 TTCTGCAGCTCGACCATGCCCAAGAGAGAGAGAGTGTCCAAATTTACTGACCGTAC 98

FEATURES

Location/Qualifiers

1..549

/organism="Apis mellifera"

/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly

A.m. ligustica"

/db_xref="taxon:7460"

RESULT 6 374 bp DNA linear GSS 05-MAR-2001

AZ892772

LOCUS

DEFINITION

RPCI-24-181M1.TJ RPCI-24 Mus musculus genomic clone

Genomic survey sequence.

ACCESSION

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:25:56 ; Search time 390.934 Seconds
(without alignments)
4888.751 Million cell updates/sec

Title: US-09-954-483B-3
Perfect score: 64
Sequence: 1 gaattccctgcagaccat.....gaaaccagtaagttatacg 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_tod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29.4	45.9	831	CNS042HQ	AL271511 Tetraodon
C 2	29.2	45.6	424	CC499557	CC499557 CH240.337
C 3	28.6	44.7	654	B6662484	B6662484 C186901.z
4	28.2	44.1	549	B1512790	B1512790 B160010A

5	28	43..8	288	29	FR0013439
6	27.6	43.1	374	28	AZ892772
7	27.6	43.1	587	9	AI667004
C 8	27.4	42.8	150	28	BH904297
C 9	27.4	42.8	380	12	B1233419
10	27.4	42.8	483	12	BI381359
11	27.4	42.8	1055	28	CC298512
12	27.4	42.8	1088	28	CC238974
13	27.2	42.5	423	10	BB678464
14	27.2	42.5	435	13	BY262497
C 15	27.2	42.5	527	29	CE334729
C 16	27.2	42.5	531	10	BE754744
C 17	27.2	42.5	643	29	CE107012
18	27.2	42.5	955	13	EU506105
19	27.2	42.5	3892	11	AK032021
C 20	27.2	42.2	309	14	CD068420
21	27.2	42.2	468	14	CD009097
22	27.2	42.2	564	14	CD714461
23	27.2	42.2	584	14	CD011119
24	27.2	42.2	586	14	CD009611
25	27.2	42.2	601	14	CD006420
26	27.2	42.2	601	14	CD716299
C 27	27.2	42.2	603	10	AW697771
28	27.2	42.2	604	14	CD716867
29	27.2	42.2	613	14	CD720661
30	27.2	42.2	614	14	CD720604
31	27.2	42.2	615	14	CD008936
32	27.2	42.2	616	14	CD010374
33	27.2	42.2	623	14	CD721317
34	27.2	42.2	630	14	CD008959
35	27.2	42.2	645	14	CD713758
36	27.2	42.2	656	14	CD010183
37	27.2	42.2	667	14	CD713853
38	27.2	42.2	669	14	CD005823
39	27.2	42.2	669	14	CD719355
40	27.2	42.2	673	14	CD720435
41	27.2	42.2	675	14	CD715328
42	27.2	42.2	675	14	CD717991
43	27.2	42.2	675	14	CD719525
44	27.2	42.2	675	14	CD721187
45	27.2	42.2	675	14	CD721208

ALIGNMENTS

RESULT 1
CNS042HQ 831 bp DNA linear GSS 01-SBP-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 077P17 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL271511
VERSION AL271511.1 GI:7993488
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissenbach J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
REFERENCE 2
AUTHORS Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C., Fizames C., Fischer C., Bouneau L., Billault A., Quetier F.,

Db 1276 ATG 1274

Best Local Similarity 67.3%; Pred. No. 1e+02;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 AATTCACTGCCAGACCATGCCAAAAAGAGAGAGAAAGGTGATGAAACCACTAAC 56
Db 3749 AATACTCCTTCTATACCACCCCTAAAAACAAAATAAACTAATAAATCAAAAAC 3695

Search completed: April 22, 2004, 13:16:14
Job time : 56.4729 secs

Db 1276 ATG 1274

Query Match 40.9%; Score 26.2; DB 17; Length 2577;
Best Local Similarity 63.5%; Pred. No. 82;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 AATTCACTGCCAGACCATGCCAAAAAGAGAGAGAAAGGTGATGAAACCACTAAC 61
Db 1750 AGTACATCTGCCAGATCAAGTGAAGAGAGTGAAGCAAACTCATGATCCTGACAAATCAA 1691

QY 62 ACG 64
Db 1690 ATG 1688

RESULT 14
US-10-389-566-242/c
; Sequence 242, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 242
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Zea mays
US-10-389-566-242

Query Match 40.9%; Score 26.2; DB 17; Length 5827;
Best Local Similarity 63.5%; Pred. No. 82;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 AATTCACTGCCAGACCATGCCAAAAAGAGAGAGAAAGGTGATGAAACCACTAAC 61
Db 1750 AGTACATCTGCCAGATCAAGTGAAGAGAGTGAAGCAAACTCATGATCCTGACAAATCAA 1691

QY 62 ACG 64
Db 1690 ATG 1688

RESULT 15
US-10-257-166-135/c
; Sequence 135, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPEROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 135
; LENGTH: 5827
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (4768..4769, 4792)
US-10-257-166-135

Query Match 40.9%; Score 26.2; DB 17; Length 5827;


```

; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101777
; LENGTH: 2420
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101777

Query Match 41.6%; Score 26.6; DB 13; Length 2420;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GCCAGACCATGCCAAAAGAGAGAAAGGTCTATGAAACCAAGTAACGTT 59
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2364 GCGAGACTCCGCTCTAAAAAAGAGAGAAAGGTCTATGAAACCAAGTAACGTT 2412
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-10-027-632-101777
; Sequence 101777, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 101777
; LENGTH: 2420
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-101777

Query Match 41.6%; Score 26.6; DB 16; Length 2420;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GCCAGACCATGCCAAAAGAGAGAAAGGTCTATGAAACCAAGTAACGTT 59
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2364 GCGAGACTCCGCTCTAAAAAAGAGAGAAAGGTCTATGAAACCAAGTAACGTT 2412
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-09-814-353-15158/c
; Sequence 15158, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15158
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-15158

Query Match 41.2%; Score 26.4; DB 10; Length 479;
Best Local Similarity 75.0%; Pred. No. 46;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 11 GCCAGACCATGCCAAAAGAGAGAAAGGTCTATGAAACCAAGTA 54
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 436 GCGAGACCTCTCTCAAGAGAGAAAGTAATCAACAGTTA 393
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-10-425-114-35487/c
; Sequence 35487, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)S
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35487
; LENGTH: 2131
; TYPE: DNA
; ORGANISM: Zea mays
US-10-425-114-35487

Query Match 40.9%; Score 26.2; DB 13; Length 2131;
Best Local Similarity 63.5%; Pred. No. 78;
Matches 40; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 AATCACCTGCCAGACCATGCCAAAAGAGAGAAAGGTCTATGAAACCAAGTAACGTTAT 61
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1336 AGTACATCTGCCAGATCAAAAGTGAAGAGTAGAGCAAACTCATGAATCCTGACAAATCAA 1277
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 62 ACG 64
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Publication No. US20040003420A1

GENERAL INFORMATION:
APPLICANT: KUEHN, Ralf
APPLICANT: FELDER, Susanne
APPLICANT: SCHWENK, Frieder
APPLICANT: KUETER-LUKS, Birgit
APPLICANT: FAUST, Nicole
TITLE OF INVENTION: Modified Recombinase
FILE REFERENCE: 012787wo/JH/ml
CURRENT APPLICATION NUMBER: US/10/014,099F
CURRENT FILING DATE: 2001-11-12
NUMBER OF SEQ ID NOS: 108
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 92
LENGTH: 70

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: primer SSV1-2

US-10-014-099F-92

Query Match 42.5%; Score 27.2; DB 16; Length 70;

Best Local Similarity 80.0%; Pred. No. 16;

Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 CTGCCAGACCATGCCCAAAAAGAGAGAAAGGTCTATGAAA 48

Db 10 CGGCCGACCATGCCCAAGAGAGAGAGAGAGAGGTGACGAAA 49

RESULT 7

US-10-014-099F-71

Sequence 71, Application US/10014099F

Publication No. US20040003420A1

GENERAL INFORMATION:

APPLICANT: KUEHN, Ralf

APPLICANT: FELDER, Susanne

APPLICANT: SCHWENK, Frieder

APPLICANT: KUETER-LUKS, Birgit

APPLICANT: FAUST, Nicole

TITLE OF INVENTION: Modified Recombinase

FILE REFERENCE: 012787wo/JH/ml

CURRENT APPLICATION NUMBER: US/10/014,099F

CURRENT FILING DATE: 2001-11-12

NUMBER OF SEQ ID NOS: 108

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 71

LENGTH: 3927

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: vector

US-10-014-099F-71

Query Match 42.5%; Score 27.2; DB 16; Length 3927;

Best Local Similarity 80.0%; Pred. No. 42;

Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 CTGCCAGACCATGCCCAAAAAGAGAGAAAGGTCTATGAAA 48

Db 670 CGGCCGACCATGCCCAAGAGAGAGAGAGAGAGGTGACGAAA 709

RESULT 8

US-10-014-099F-75

Sequence 75, Application US/10014099F

Publication No. US20040003420A1

GENERAL INFORMATION:

APPLICANT: KUEHN, Ralf

APPLICANT: FELDER, Susanne

APPLICANT: SCHWENK, Frieder

APPLICANT: KUETER-LUKS, Birgit

APPLICANT: FAUST, Nicole
TITLE OF INVENTION: Modified Recombinase
FILE REFERENCE: 012787wo/JH/ml
CURRENT APPLICATION NUMBER: US/10/014,099F
CURRENT FILING DATE: 2001-11-12
NUMBER OF SEQ ID NOS: 108
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 75

LENGTH: 4905

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: vector

US-10-014-099F-75

Query Match 42.5%; Score 27.2; DB 16; Length 4905;

Best Local Similarity 80.0%; Pred. No. 45;

Matches 32; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 9 CTGCCAGACCATGCCCAAAAAGAGAGAAAGGTCTATGAAA 48

Db 966 CGGCCGACCATGCCCAAGAGAGAGAGAGAGGTGACGAAA 1005

RESULT 9

US-10-424-599-129852/c

Sequence 129852, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 129852

LENGTH: 393

TYPE: DNA

ORGANISM: Glycine max

FEATURE:

OTHER INFORMATION: Clone ID: PAT_MRT3847_88265C.1

US-10-424-599-129852

Query Match 41.6%; Score 26.6; DB 13; Length 393;

Best Local Similarity 66.7%; Pred. No. 38;

Matches 38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Qy 4 TTCACCTGCCAGACCATGCCCAAAAAGAGAGAAAGGTCTATGAAAACGTAACGTTA 60

Db 63 TTCCCTGCCCGCTCTGCTTAAAGTTAAAGAGGTCGCGAATCAGGAACGGTA 7

RESULT 10

US-10-027-632-101777

Sequence 101777, Application US/10027632

Publication No. US20020198371A1

GENERAL INFORMATION:

APPLICANT: Wang David G

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

```
RESULT 2
US-09-954-483A-4
; Sequence 4, Application US/09954483A
; Publication No. US20030032175A1
; GENERAL INFORMATION:
; APPLICANT: Siebel, Christian
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF PRODUCING CELLS AND ANIMALS
; TITLE OF INVENTION: COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: RMES-02
; CURRENT APPLICATION NUMBER: US/09/954,483A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/232,957
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 66
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-954-483A-4

Query Match      100.0%; Score 64; DB 10; Length 66;
Best Local Similarity 100.0%; Pred. No. 7.9e-12;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTCCGACGATGCCAAAAGAGAGAAAGGTCATGAACCGTAACGTTA 60
Db 3 GAATTCACCTCCGACGATGCCAAAAGAGAGAAAGGTCATGAACCGTAACGTTA 62
QY 61 TACG 64
Db 63 TACG 66

RESULT 3
US-09-954-483A-13
; Sequence 13, Application US/09954483A
; Publication No. US20030032175A1
; GENERAL INFORMATION:
; APPLICANT: Siebel, Christian
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF PRODUCING CELLS AND ANIMALS
; TITLE OF INVENTION: COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: RMES-02
; CURRENT APPLICATION NUMBER: US/09/954,483A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/232,957
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 6148
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Construct Sequence
US-09-954-483A-13

Query Match      100.0%; Score 64; DB 10; Length 6148;
Best Local Similarity 100.0%; Pred. No. 2.5e-11;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTCCGACGATGCCAAAAGAGAGAAAGGTCATGAACCGTAACGTTA 60
Db 4213 GAATTCACCTCCGACGATGCCAAAAGAGAGAAAGGTCATGAACCGTAACGTTA 4272
QY 61 TACG 64.
```

```
Db 4273 TACG 4276

RESULT 4
US-09-872-868-17/c
; Sequence 17, Application US/09872868
; Publication No. US20030166191A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Timothy
; TITLE OF INVENTION: Bistable Genetic Toggle Switch
; FILE REFERENCE: CEL-002
; CURRENT APPLICATION NUMBER: US/09/872,868
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: PCT/US99/28592
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: US 60/110,616
; PRIOR FILING DATE: 1998-12-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 5525
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pIKE107
US-09-872-868-17

Query Match      51.2%; Score 32.8; DB 10; Length 5525;
Best Local Similarity 76.9%; Pred. No. 0.62;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 CAGACCATGCGCAAAAAGAGAGAAAGGTCATGAACCGTAACGTTATACG 64
Db 5438 CTGACCAGGAGGAAACAGACCATGCGCATGAACCGTAACGTTATACG 5387

RESULT 5
US-09-872-868-16/c
; Sequence 16, Application US/09872868
; Publication No. US20030166191A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Timothy
; TITLE OF INVENTION: Bistable Genetic Toggle Switch
; FILE REFERENCE: CEL-002
; CURRENT APPLICATION NUMBER: US/09/872,868
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: PCT/US99/28592
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: US 60/110,616
; PRIOR FILING DATE: 1998-12-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 5522
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid pIKE105
US-09-872-868-16

Query Match      42.8%; Score 27.4; DB 10; Length 5522;
Best Local Similarity 83.8%; Pred. No. 40;
Matches 31; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 28 AAGAGAGAAAGGTCATGAACCGTAACGTTATACG 64
Db 5423 AAAAATGAGCGTCATGAACCGTAACGTTATACG 5387

RESULT 6
US-10-014-099F-92
; Sequence 92, Application US/10014099F
```

Result No.	Query			ID	Description
	Score	Match	Length		
1	64	100.0	64	US-09-954-483A-3	Sequence 3, Appli
2	64	100.0	66	US-09-954-483A-4	Sequence 4, Appli
3	64	100.0	6148	US-09-954-483A-13	Sequence 13, Appli
C 4	32.8	51.2	5525	US-09-872-868-17	Sequence 16, Appli
C 5	27.4	42.8	5522	US-09-872-868-16	Sequence 17, Appli
6	27.2	42.5	70	US-10-014-099F-92	Sequence 92, Appli
7	27.2	42.5	3927	US-10-014-099F-71	Sequence 71, Appli
8	27.2	42.5	4905	US-10-014-099F-75	Sequence 75, Appli
C 9	26.6	41.6	393	US-10-424-599-129852	Sequence 129852
10	26.6	41.6	2420	US-10-027-632-101777	Sequence 101777
C 11	26.6	41.6	2420	US-10-027-632-101777	Sequence 101777
C 12	26.4	41.2	479	US-09-814-353-15158	Sequence 15158, A
C 13	26.2	40.9	2131	US-10-425-114-35487	Sequence 35487, A
C 14	26.2	40.9	2577	US-10-389-566-242	Sequence 242, App

```

; LENGTH: 792
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2275

Query Match      36.9%; Score 23.6; DB 4; Length 792;
Best Local Similarity 69.6%; Pred. No. 34;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      18 CATGCCAAAAAGAGAGAAAGTTCATGAACCCAGTAAAGTTATAC 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      511 CATCCAAAAAAGATCCGAAAGCTTATAAAGCAGAGATGCAATGC 556

```

Search completed: April 22, 2004, 12:58:16
Job time : 14.3382 secs


```

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; FILE REFERENCE: jannaschii
; CURRENT APPLICATION NUMBER: US/09/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191589)..(191589)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319236)..(319236)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)

```

GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
Amalfitano, Andrea
Hauser, Michael A.
Kumar-Singh, Rajendra
Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/562,919
FILING DATE: 02-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE: 23-Oct-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-562-919-11
Query Match 39.1%; Score 25; DB 4; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 15 GACCATCCCAAAAGGAAGGAAGGTCATGAAACCAAGTAACTTATAC 63
Db 314 GACCATCCCAAAAGGAAGGAAGGTTCCCAATTACTGACCGTACAC 362
RESULT 9
US-08-217-360-9
Sequence 9, Application US/08217360
Patent No. 5530191
GENERAL INFORMATION:
APPLICANT: MALIGA, Pal
TITLE OF INVENTION: METHODS FOR PRODUCING CYTOPLASMIC
TITLE OF INVENTION: MALE STERILITY IN PLANTS AND USE THEREOF IN PRODUCTION OF HYBR
TITLE OF INVENTION: SEED
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman,
ADDRESS: P.C.
STREET: 1601 Market Street, Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,360
FILING DATE: 24-MAR-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: REED, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: Rutgers University
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215)563-4100
TELEFAX: (215)563-4044
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-217-360-9
Query Match 37.8%; Score 24.2; DB 1; Length 240;
Best Local Similarity 89.7%; Pred. No. 15;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 36 AAAGTCATGAACCAACGTAACGTTATACG 64
Db 189 AAGTGCATGCACCAACGTAACGTTATACG 217
RESULT 10
US-09-621-976-3126/c
Sequence 3126, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 3126
LENGTH: 707
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 12..203
US-09-621-976-3126
Query Match 37.2%; Score 23.8; DB 4; Length 707;
Best Local Similarity 62.7%; Pred. No. 28;
Matches 37; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 4 TTCACCTGCCAGACCATGCCCAAAAGGAAGGTCATGAACCAAGTAACTTATA 62
Db 531 TTCCAATTCCAAACCTCACAAAGAACAAAAGAAAGTCTCTGCAACATCAGCTTATA 473
RESULT 11
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.


```
; APPLICATION NUMBER: US 08/735,609
; FILING DATE: 08/735,609
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-315-372-11

Query Match 39.1%; Score 25; DB 3; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 15 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 63
Db 314 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 362

RESULT 6
US-09-244-752-11
; Sequence 11, Application US/09244752
; Patent No. 6063622
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/244,752
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/735,609
; APPLICATION NUMBER: <B> FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-244-752-11

Query Match 39.1%; Score 25; DB 3; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 15 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 63
Db 314 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 362

RESULT 7
US-09-245-497-11
; Sequence 11, Application US/09245497
; Patent No. 6083750
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/245,497
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/735,609
; APPLICATION NUMBER: <B> FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-09-245-497-11

Query Match 39.1%; Score 25; DB 3; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 15 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 63
Db 314 GACCATGCCAAAGAGAGAGAGTGCATGAAACCGTAGCGTTATAC 362

RESULT 8
US-09-562-919-11
; Sequence 11, Application US/09562919
; Patent No. 6451596
```

Best Local Similarity 64.9%; Pred. No. 8.9; Mismatches 20; Indels 0; Gaps 0;
Matches 37; Conservative 0;

QY 2 AATTCACCTGCCAGACCCGCAAAAAGAGAAAGGTCATGAACCAAGTAACTG 58
DB 254 AATTCCTCTCAATTGCATGTGTAAGAGAGATAATCTTCATGAGATAAAAAATGT 198

RESULT 3

US-08-735-609-11
; Sequence 11, Application US/08735609
; Patent No. 595360
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/735,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02484
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-735-609-11

Query Match 39.1%; Score 25; DB 2; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 15 GACCATGCCCAAGAGAAAGGTCATGAACCAAGTAACTGATAC 63
DB 314 GACCATGCCCAAGAGAAAGGTCATGAACCAAGTAACTGATAC 362

RESULT 4

US-08-735-609-11
; Sequence 11, Application US/08735609
; Patent No. 5994132
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE: 23-Oct-1996
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-08-735-609-11

Query Match 39.1%; Score 25; DB 2; Length 699;
Best Local Similarity 69.4%; Pred. No. 11;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 15 GACCATGCCCAAGAGAAAGGTCATGAACCAAGTAACTGATAC 63
DB 314 GACCATGCCCAAGAGAAAGGTCATGAACCAAGTAACTGATAC 362

RESULT 5

US-09-315-372-11
; Sequence 11, Application US/09315372
; Patent No. 6057158
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,372
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:26:21 ; Search time 11.3382 Seconds
(without alignments)
3132.491 Million cell updates/sec

Title: US-09-954-483B-3
Perfect score: 64
Sequence: 1 gaattcaactgcccagaccat.....gaaaccagtaacgttatatcg 64

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	40.6	852	US-09-328-352-905	Sequence 905, App
2	25	39.1	348	US-09-134-001C-2127	Sequence 2127, Ap
3	25	39.1	699	US-08-735-609-11	Sequence 11, Appl
4	25	39.1	699	US-08-735-609-11	Sequence 11, Appl
5	25	39.1	699	US-09-315-372-11	Sequence 11, Appl
6	25	39.1	699	US-09-244-752-11	Sequence 11, Appl
7	25	39.1	699	US-09-245-497-11	Sequence 11, Appl
8	25	39.1	699	US-09-562-919-11	Sequence 11, Appl
9	24.2	37.8	240	US-08-217-360-9	Sequence 9, Appl
10	23.8	37.2	707	US-09-621-976-3126	Sequence 3126, Ap
11	23.8	37.2	1664976	US-08-916-421B-1	Sequence 1, Appl
12	23.6	36.3	197	US-09-872-761A-10	Sequence 10, Appl
13	23.6	36.9	205	US-09-872-761A-1	Sequence 1, Appl
14	23.6	36.9	515	US-09-621-976-11502	Sequence 11502, A
15	23.6	36.9	792	US-09-134-001C-2275	Sequence 2275, Ap
16	23.6	36.9	1353	US-08-956-171E-521	Sequence 521, App
17	23.6	36.9	1701	US-09-543-681A-372	Sequence 372, App
18	23.4	36.6	36	US-08-392-771-29	Sequence 29, Appl
19	23.4	36.6	37	US-08-392-771-24	Sequence 24, Appl
20	23.4	36.6	2509	US-08-954-333-6	Sequence 6, Appl
21	23.4	36.6	2797	US-08-954-333-8	Sequence 8, Appl
22	23.4	36.6	5578	US-08-081-610-2	Sequence 2, Appl
23	23.2	36.2	1404	US-09-107-532A-2521	Sequence 2521, Ap
24	23.2	36.2	2100	US-08-938-830-2	Sequence 2, Appl
25	23.2	36.2	2100	US-09-020-222-2	Sequence 2, Appl
26	23	35.9	360	US-09-540-236-1324	Sequence 1324, Ap
27	23	35.9	1500	US-09-134-000C-2664	Sequence 2664, Ap

```

28      23      35.9      3150      4      US-09-522-666-5      Sequence 5, Appli
c 29      23      35.9      4307      4      US-09-803-671B-1      Sequence 1, Appli
30      23      35.9      9542      3      US-08-968-685A-9      Sequence 9, Appli
31      23      35.9      62909      4      US-09-596-002-32      Sequence 32, Appli
32      23      35.9      64667      4      US-09-803-671B-3      Sequence 3, Appli
33      22.8      35.6      101      3      US-09-411-687A-25      Sequence 25, Appli
34      22.8      35.6      230      4      US-09-702-705-533      Sequence 533, App
35      22.8      35.6      290      4      US-09-736-457-533      Sequence 533, App
36      22.8      35.6      230      4      US-09-614-124B-533      Sequence 533, App
37      22.8      35.6      230      4      US-09-671-325-533      Sequence 533, App
38      22.8      35.6      290      4      US-09-589-184-533      Sequence 533, App
39      22.8      35.6      1100      2      US-08-886-633-1      Sequence 1, Appli
40      22.8      35.6      1100      3      US-09-213-081-1      Sequence 1, Appli
41      22.8      35.6      1100      3      US-09-212-979-1      Sequence 1, Appli
42      22.8      35.6      1239      3      US-09-411-687A-27      Sequence 27, Appli
c 43      22.8      35.6      2994      4      US-09-549-872B-7      Sequence 7, Appli
44      22.8      35.6      5261      4      US-09-770-315-7      Sequence 7, Appli
c 45      22.8      35.6      392000      4      US-10-027-983-11      Sequence 11, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-328-352-905
; Sequence 905, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 905
; LENGTH: 852
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-905

```

```

Query Match      40.6%; Score 26; DB 4; Length 852;
Best Local Similarity 70.0%; Pred.No. 5.2;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

```

Cy 1 GAATTCACTGCCAGACCATGCCAAAAAGAGAGAGAGAGTCTATGAAC 50
Db 2 GAGTAACTGATGGACCATGAACCTAGACAGAGAGAGTCTATGAAC 51

```

```

RESULT 2
US-09-134-001C-2127/c
; Sequence 2127, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2127
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2127

```

```

Query Match      39.1%; Score 25; DB 4; Length 348;

```

Job time : 58.6603 secs

```

RESULT 15
AAZ07752
ID AAZ07752 standard; DNA; 61 BP.
XX
AC AAZ07752;
XX
DT 23-NOV-1999 (first entry)
XX
DE Human P450 reductase derivative alp450R amplifying 5' primer.
XX
KW Prodrug; localization domain; tumor-selective antibody; cytochrome P450;
KW prodrug activating domain; modified hematopoietic stem cell; MHC; tumor;
KW inflammation; atherosclerosis; muscular dystrophy; cerebral malaria;
KW rheumatoid arthritis; hypoxia; ischemia; hypoglycemia; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9945126-A2.
XX
FD 10-SEP-1999.
XX
PF 05-MAR-1999; 99WO-GB000672.
XX
PR 06-MAR-1998; 98GB-00004841.
PR 19-AUG-1998; 98GB-00018103.
PR 29-JAN-1999; 99GB-00002081.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;
XX PI Mitrophanous K;
XX
XX WPI; 1999-540852/45.
XX
XX New prodrug activating agent targeted to selected cells or tissues.
XX PT Particularly hypoxic cells, for treating e.g. tumors or inflammation.
XX
XX Example 1; Page 80; 149pp; English.
XX
XX The invention provides a new prodrug activating agent that comprises: (i)
XX a localization domain (LD; other than a tumor-selective antibody) and a
XX prodrug activating domain (PAD); (ii) at least one nucleic acid encoding
XX a cytochrome P450 and under control of at least one constitutive or
XX inducible expression control sequence or (iii) a modified hematopoietic
XX stem cell (MHC) containing at least one nucleic acid encoding a PAD and
XX under control of elements as in (ii). The prodrug activating agent or
XX vectors that express them, are specifically used to treat tumors,
XX inflammation, atherosclerosis and muscular dystrophy, but may also be
XX used to treat many other conditions, e.g. cerebral malaria, rheumatoid
XX arthritis, or conditions associated with hypoxia, hypoglycemia or
XX ischemia, or to deliver antibiotics, antiviral agents, analgesics,
XX anesthetics, anti-inflammatories, antineoplastic agents and diagnostic
XX agents. ID optimize activity of PAD, e.g. by delivering it to selected
XX locations or by delivering it to neighboring cells (bystander effect),
XX and allow a reduction in dose of prodrug, and thus of systemic side-
XX effects. Nucleic acids encoding the agent may be expressed selectively in
XX hypoxic cells. Sequences AAZ07752-53 represent primers for amplifying
XX derivatives of human P450 reductase
XX
SQ Sequence 61 BP; 20 A; 17 C; 15 G; 9 T; 0 U; 0 Other;
Query Match 40.6%; Score 26; DB 2; Length 61;
Best Local Similarity 100.0%; Pred No. 33;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 ACCATGCCAAAAGAGAGAGAGGT 41
Db 10 ACCATGCCAAAAGAGAGAGAGGT 35

```

Search completed: April 22, 2004, 10:42:33

Qy 16 ACCATGCCAAAAAGAGAAAGGT 41
Db 9 ACCATGCCAAAAAGAGAAAGGT 34

RESULT 13
AAZ07756
ID AAZ07756 standard; DNA; 60 BP.

Db 9 ACCATGCCAAAAAGAGAGAAAGGT 34

RESULT 14
AAZ19789
ID AAZ19789 standard; DNA; 61 BP.

CC into the genomes of the cells. (I) is capable of modifying target gene in
 CC a cell with high efficiency and specificity. Use of (I) provides a faster
 CC and more efficient means for isolating and selecting cells comprising
 CC target gene modification. Also use of (I) provides an increase over
 CC previous technologies in both the speed and frequency at which homologous
 CC recombination events can be recovered. (I) is also useful for creation of
 CC transgenic animals containing targetted gene modifications. This sequence
 CC represents the positive selection vector construct c3406

XX Sequence 6148 BP; 1354 A; 1593 C; 1770 G; 1431 T; 0 U; 0 Other;

Query Match 100.0%; Score 64; DB 6; Length 6148;
 Best Local Similarity 100.0%; Pred. No. 1.7e-11;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTGCGAGACCATGCGCAAAAAGAGAGAGAGAGTTCATGAAACCCAGTAACTGTTA 60
 DB 4213 GAATTCACCTGCGAGACCATGCGCAAAAAGAGAGAGAGTTCATGAAACCCAGTAACTGTTA 4272

QY 61 TAGC 54

DB 4273 TAGC 4276

RESULT 4

ID AAZS0628 standard; DNA; 9407 BP.

XX AAZS0628;

XX 06-AUG-2003 (revised)

XX 23-MAY-2000 (first entry)

XX Recombinant plasmid pMPG3.

XX Adenoviral vector; lac repressor; rabies glycoprotein G; plasmid pMPG3;
 XX lac operator; human cytomegalovirus promoter; HCMV; LAP348; ds.

XX Synthetic.

XX Human adenovirus type 5.

XX Human herpesvirus 5.

XX Simian virus 40.

XX Human herpesvirus 1.

XX WO200006751-A2.

XX 10-FEB-2000.

XX 30-JUL-1999; 99WO-IB001349.

XX 30-JUL-1998; 98US-0094681P.

XX 31-DEC-1998; 98US-00223820.

XX (ADVE-) ADVEC INC.

XX Graham FL, Prevec L, Matthews DA;

XX WPI; 2000-195309/17.

XX New recombinant eukaryotic cells, containing a polynucleotide encoding a
 XX repressor protein and an expression vector that produces a heterologous
 XX protein.

XX Example 1; Fig 8; 66pp; English.

XX The patent discloses a method of propagating recombinant adenoviral
 CC vectors in eukaryotic cells expressing lac repressor protein. The vector
 CC comprises of nucleic acid encoding a heterologous protein (like rabies
 CC glycoprotein G) linked to a lac operator region, to which the lac
 CC repressor binds and represses the expression of heterologous protein. The
 CC eukaryotic cell lines can be used for replicating recombinant adenoviral
 CC vectors by inhibiting production of certain viral-proteins whose
 CC overexpression inhibits the production of viruses. The present sequence

CC is recombinant plasmid pMPG3. It comprises of E1 region- deleted
 CC adenoviral genome, lac repressor (LAP348) gene, Simian virus 40 derived
 CC Poly A addition sequence, transcription activation domain from Herpes
 CC simplex virus type 1 VP16, human cytomegalovirus (HCMV) promoter and a
 CC copy of adenovirus inverted terminal repeats (ITRs). These plasmids can
 CC be used to transfect eukaryotic cells. (Updated on 06-AUG-2003 to correct
 CC OS field.)

XX Sequence 9407 BP; 2094 A; 2462 C; 2568 G; 2283 T; 0 U; 0 Other;

Query Match 53.8%; Score 34.4; DB 3; Length 9407;
 Best Local Similarity 52.3%; Pred. No. 0.15;
 Matches 48; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 16 ACCATGCGCAAAAAGAGAGAGAGT---CATGAAACCCAGTAACTGTTATACG 64

DB 735 ATCATGCGCAAAAAGAGAGAGAGTAAACATGAAACCCAGTAACTGTTATACG 786

RESULT 5

AAI90264

ID AAI90264 standard; cDNA; 419 BP.

XX AAI90264;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 10324.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US004927.

XX 28-FEB-2000; 2000US-00515126.

XX 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX P-PSDB; AAO10333.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 XX and treating e.g. leukemia, inflammation and immune disorders.

XX Claim 1; SEQ ID NO 10324; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 419 BP; 117 A; 94 C; 113 G; 95 T; 0 U; 0 Other;

CC cells having targeting vector integrated via homologous recombination
 CC into the genomes of the cells. (I) is capable of modifying target gene in
 CC a cell with high efficiency and specificity. Use of (I) provides a faster
 CC and more efficient means for isolating and selecting cells comprising
 CC target gene modification. Also use of (I) provides an increase over
 CC previous technologies in both the speed and frequency at which homologous
 CC recombination events can be recovered. (I) is also useful for creation of
 CC transgenic animals containing targeted gene modifications. This sequence
 CC represents the SV40 large T antigen nuclear localisation signal
 CC incorporated into the positive selection vector constructs described in
 CC the invention

XX Sequence 64 BP; 27 A; 14 C; 13 G; 10 T; 0 U; 0 Other;

Query Match 100.0%; Score 64; DB 6; Length 64;
 Best Local Similarity 100.0%; Pred. No. 6.7e-12;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCAAGTAACGTTA 60
 |||||
 Db 1 GAATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCAAGTAACGTTA 60
 |||||
 QY 61 TACG 64
 |||||
 Db 61 TACG 64

RESULT 2
 ABK49511
 ID ABK49511 standard; DNA; 66 BP.

XX AC ABK49511;

DT 15-JUL-2002 (first entry)

XX Positive selection vector associated oligonucleotide 10164.

XX Transgenic animal; targeting vector; positive selection vector;
 KW homologous recombination; target gene modification; transgenic animal;
 KW ss.

XX Synthetic.

XX WO200222834-A2.

XX 21-MAR-2002.

XX 17-SEP-2001; 2001WO-US028892.

XX 15-SEP-2000; 2000US-0232957P.

XX (DELT-) DELTAGEN INC.

XX Siebel C, Brennan TJ;

XX WPI; 2002-383132/41.

XX Novel targeting vector modifying target gene, has first and second
 PT sequences homologous to target gene portions, a selectable marker
 PT cassette and regulator, useful for producing animals with targeted gene
 PT modifications.

XX Example 1; Fig 8; 43pp; English.

XX The invention describes a targeting vector (positive selection vector)
 CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
 CC and S2) homologous to a portion or region of a target gene, a selectable
 CC marker cassette and a regulator. (I) is useful for producing cells
 CC comprising a modification of the target gene which involves introducing
 CC (I) into cells capable of homologous recombination, selecting for cells
 CC expressing the selectable marker and identifying cells containing the
 CC modification of the target gene. Use of (I) for enriching cells
 CC comprising disruption or modification of target gene enhances recovery of

CC cells having targeting vector integrated via homologous recombination
 CC into the genomes of the cells. (I) is capable of modifying target gene in
 CC a cell with high efficiency and specificity. Use of (I) provides a faster
 CC and more efficient means for isolating and selecting cells comprising
 CC target gene modification. Also use of (I) provides an increase over
 CC previous technologies in both the speed and frequency at which homologous
 CC recombination events can be recovered. (I) is also useful for creation of
 CC transgenic animals containing targeted gene modifications. This sequence
 CC represents an oligonucleotide used in the creation of the positive
 CC selection vectors for targeted gene modification

XX Sequence 66 BP; 27 A; 15 C; 14 G; 10 T; 0 U; 0 Other;

Query Match 100.0%; Score 64; DB 6; Length 66;
 Best Local Similarity 100.0%; Pred. No. 6.8e-12;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCAAGTAACGTTA 60
 |||||
 Db 3 GAATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCAAGTAACGTTA 62
 |||||
 QY 61 TACG 64
 |||||
 Db 63 TACG 66

RESULT 3
 ABK49520
 ID ABK49520 standard; DNA; 6148 BP.

XX AC ABK49520;

XX 15-JUL-2002 (first entry)

XX Positive selection vector construct c3406.

XX Transgenic animal; targeting vector; positive selection vector;
 KW homologous recombination; target gene modification; transgenic animal;
 KW c3406; ds.

XX Synthetic.

XX WO200222834-A2.

XX 21-MAR-2002.

XX 17-SEP-2001; 2001WO-US028892.

XX 15-SEP-2000; 2000US-0232957P.

XX (DELT-) DELTAGEN INC.

XX Siebel C, Brennan TJ;

XX WPI; 2002-383132/41.

XX Novel targeting vector modifying target gene, has first and second
 PT sequences homologous to target gene portions, a selectable marker
 PT cassette and regulator, useful for producing animals with targeted gene
 PT modifications.

XX Claim 32; Fig 6A; 43pp; English.

XX The invention describes a targeting vector (positive selection vector)
 CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
 CC and S2) homologous to a portion or region of a target gene, a selectable
 CC marker cassette and a regulator. (I) is useful for producing cells
 CC comprising a modification of the target gene which involves introducing
 CC (I) into cells capable of homologous recombination, selecting for cells
 CC expressing the selectable marker and identifying cells containing the
 CC modification of the target gene. Use of (I) for enriching cells
 CC comprising disruption or modification of target gene enhances recovery of
 CC cells having targeting vector integrated via homologous recombination

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:24:05 ; Search time 55.6603 Seconds
(without alignments)
4884.713 Million cell updates/sec

Title: US-09-954-483B-3
Perfect score: 64
Sequence: 1 gaattcacctgcccagaccat.....Gaaacacgtaacgttatcacg 64

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001as:.*
5: Geneseq2001bs:.*
6: Geneseq2002s:.*
7: Geneseq2003as:.*
8: Geneseq2003bs:.*
9: Geneseq2003cs:.*
10: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	64	6	Abk49510 SV40 larg
2	64	100.0	66	6	Abk49511 Positive
3	64	100.0	66	6	Abk49520 Positive
4	34.4	53.8	9407	3	Aaz50628 Recombina
5	27.4	42.8	419	4	Aai50264 Human pol
6	27.2	42.5	70	6	Abt08187 Recombina
7	27.2	42.5	3927	6	Abt08166 Recombina
8	27.2	42.5	4905	6	Abt08170 Recombina
9	26.6	41.6	944	9	Abt08170 Recombina
10	26.2	40.9	5827	6	Abk40053 Human che
11	26.2	40.9	110000	6	Continuation (17 o
12	26	40.6	60	2	Aaz19791 SV40 nucl
13	26	40.6	60	2	Aaz07756 Human P45
14	26	40.6	61	2	Aaz19789 SV40 nucl
15	26	40.6	61	2	Aaz07752 Human P45
16	26	40.6	852	8	Ada29618 DNA encod
17	26	40.6	1455	6	Aas18563 Xylella f
18	26	40.6	8129	4	Aas46763 Tumour su
19	26	40.6	12709	6	Aas18570 Xylella f
20	25.8	40.3	122748	6	Abt10719 Human bre
21	25.4	39.7	2074	9	Ades4795 Rat gene
22	25.2	39.4	378	4	Aai181038 Human pol
23	25.2	39.4	402	5	Abv16946 Human pro

24	25.2	39.4	457	5	ABV46741	Abv46741 Human pro
25	25.2	39.4	2000	6	ABZ16775	Abz16775 Arabidops
26	25.2	39.4	2293	9	ADB62992	ADB62992 Human cDN
27	25.2	39.4	2651	5	AAS72689	Aas72689 DNA encod
28	25.2	39.4	2736	9	AD862417	AD862417 Human cDN
29	25.2	39.4	3898	9	AD862450	AD862450 Human cDN
30	25	39.1	348	6	ABN92664	ABN92664 Staphyloc
31	25	39.1	699	2	AAV07268	AAV07268 Plasmid p
32	25	39.1	2573	9	ADD48932	ADD48932 Human gen
33	25	39.1	2573	9	ADD47983	ADD47983 Human gen
34	25	39.1	3170	4	AAH54485	AAH54485 S. epider
35	25	39.1	4329	4	ACA46888	ACA46888 Prokaryot
36	25	39.1	4329	4	AAH53329	AAH53329 S. epider
37	25	39.1	4847	4	AAU09280	AAU09280 PKC-cre-p
38	25	39.1	4960	4	AAU09269	AAU09269 PCMV-I-Cr
39	25	39.1	4960	6	ABT08148	ABT08148 Recombina
40	25	39.1	5878	6	ABT08199	ABT08199 Recombina
41	24.8	38.8	503	8	ACH13011	ACH13011 Human adu
42	24.8	38.8	268685	6	ABS56563	ABS56563 Human SUL
43	24.6	38.4	428	4	AAI88208	AAI88208 Human pol
44	24.6	38.4	732	5	AAS91538	Aas91538 DNA encod
45	24.6	38.4	3056	4	ABL13162	ABL13162 Drosophil

ALIGNMENTS

RESULT 1

ABK49510
ID ABK49510 standard; DNA; 64 BP.

XX ABK49510;

XX 15-JUL-2002 (first entry)

DE SV40 large T antigen nuclear localisation signal.

Transgenic animal; targeting vector; positive selection vector;

homologous recombination; target gene modification; transgenic animal;
simian virus 40; SV40; nuclear localisation signal; NLS; ds.

Rhesus macaque polyoma virus.

WO200222834-A2.

21-MAR-2002.

17-SEP-2001; 2001WO-US028892.

15-SEP-2000; 2000US-0232957P.

(DELT-) DELTAGEN INC.

Siebel C, Brennan TJ;

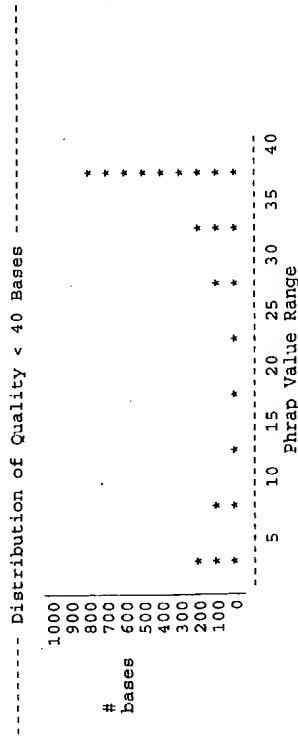
WPI; 2002-383132/41.

Novel targeting vector modifying target gene, has first and second
sequences homologous to target gene portions, a selectable marker
cassette and regulator, useful for producing animals with targeted gene
modifications.

Claim 15; Fig 6E; 43pp; English.

The invention describes a targeting vector (positive selection vector)
(i) capable of modifying a target gene. (i) comprises two sequences (S1
and S2) homologous to a portion or region of a target gene, a selectable
marker cassette and a regulator. (i) is useful for producing cells
comprising a modification of the target gene which involves introducing
(i) into cells capable of homologous recombination, selecting for cells
expressing the selectable marker and identifying cells containing the
modification of the target gene. Use of (i) for enriching cells
comprising disruption or modification of target gene enhances recovery of

130939 cagaatttca(n)mggggctgg cagaatttca(a)aggggctgg
130940 agaatttcann(n)nggggctgga agaatttcaa(a)ggggctgga
130941 gaatttcaann(n)ggggctgga gaatttcaa(a)ggggctgga
130993 ttataaaa(n)ttatacctga ttataaaa(a)ttatacctga
187546 ttgggaatt(n)c. ttgggaatt(c)c.



Version: 1.01 qkfo.
Location/Qualifiers
1. .187547
source /organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Query Match 44.4%; Score 28.4; DB 9; Length 187547;
Best Local Similarity 70.4%; Pred. NO. 1.6e+02;
Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 GAATTCACCTGCCAGACCATGCCAAAAAGAGAGAGAGGTCATGAACCCAGTA 54
Db 62133 GAATTCACCTGCATATGATCGCAGAGATATAAGAGAGATTATTAATCTGTA 62186

Search completed: April 22, 2004, 11:46:23
Job time : 363.763 secs

Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honei, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, C., Liu, W., Loulseghe, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodargren, G., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, F., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D. and Gibbs, R.

TITLE
Direct Submission
REFERENCE
Unpublished
2 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (13-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (28-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (29-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (07-APR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (01-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
8 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (02-MAY-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
9 (bases 1 to 187547)
AUTHORS
Worley, K.C.

AUTHORS
Worley, K.C.
TITLE
Direct Submission
REFERENCE
Submitted (21-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
10 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
11 (bases 1 to 187547)
AUTHORS
Worley, K.C.

TITLE
Direct Submission
REFERENCE
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 23, 2001 this sequence version replaced gi:13123832.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
(Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URU:
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

----- Summary Statistics -----	
Contig length:	187547
Phrap values in estimate:	182369
Average error rate (BCM-Phrap estimate):	9.09188e-06
Fraction of Phrap values less than 40 :	0.00664038
Number of consensus changing edits:	9
Number of N's in consensus :	0

----- Consensus changing edits -----	
Position	Original+Context
41688	tctctttca(n)atacaaaaa
85086	cagaaggttg(n)tttaaatca
94830	aaaaaaaaa(a)cagaaaaaa
94850	agaaaaagaa(a)ctgtgctgt

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES

```

Source
Location/Qualifiers
1. .171004
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-105E24"
3131. .3167
/rpt_family="CA)n"
complement(4089. .4156)
/rpt_family="MER103"
4791. .4814
/rpt_family="AT_rich"
5319. .5582
/rpt_family="AluJo"
complement(5984. .6414)
/rpt_family="MLT1K"
6506. .6819
/standard_name="103866"
complement(6584. .6694)
/rpt_family="MLT1K"
7004. .7091
/rpt_family="(TA)n"
complement(7115. .7699)
/rpt_family="MLT2D"
complement(7803. .8329)
/rpt_family="MLT1E1"
10180. .10482
/rpt_family="AluSc"
complement(10900. .11246)
/rpt_family="MLT1H"
11305. .11338
/rpt_family="AT_rich"
12649. .12804
/rpt_family="MIR"
12844. .12940
/rpt_family="MIR"
complement(13008. .13326)
/rpt_family="L2"
complement(14450. .14564)
/rpt_family="MIR"
complement(15887. .16019)
/rpt_family="LIMA4"
complement(16652. .16748)
/rpt_family="MER5A"
16755. .16841
/rpt_family="(TTTC)n"
16857. .16954
/rpt_family="CT_rich"
16957. .16988
/rpt_family="MER5A"
17504. .17718
/rpt_family="L2"
17755. .17836

```

```

/rpt_family="MIR"
complement(19183. .19453)
/rpt_family="MER58B"
complement(19584. .19998)
/rpt_family="MSTB"
19971. .20230
/standard_name="73085"
21270. .21836
/rpt_family="MLT1E2"
21940. .22237
/rpt_family="AluJo"
22240. .22548
/rpt_family="MER2"
complement(22549. .22593)
/rpt_family="L2"
complement(22722. .23557)
/rpt_family="L2"
23729. .23800
/rpt_family="(GA)n"
23872. .23892
/rpt_family="AT_rich"
24050. .24071
/rpt_family="AT_rich"
complement(24208. .24297)
/rpt_family="MIR"
complement(24303. .24803)
/rpt_family="MLT2CB"
25345. .25393
/rpt_family="AT_rich"
25454. .25548
/rpt_family="MIR"
26227. .26277
/rpt_family="A-rich"
complement(27551. .27764)
/rpt_family="MIR"

```

Query Match 44.4%; Score 28.4; DB 9; Length 171004;

Best Local Similarity 66.1%; Pred. No. 1.6e+02; Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

QY 2 AATTCACCTGCCAGACCATGCCCAAAAAGAGAGAAAGGTCTATGAACAGTAACCTTAT 61
Db 89681 AATCAGTAGCCAGAACTTCACAAAGCTGATGATTAAGTATCATGATATCATTA 89740
QY 62 AC 63
Db 89741 AC 89742

```

RESULT 15

```

AC078958 AC078958 187547 bp DNA linear PRI 02-MAY-2002
LOCUS Homo sapiens X BAC RP11-609C15 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC078958
VERSION AC078958.30 GI:13435186
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187547)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Albretrooks,S.L., Amaraturge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brivea,M., Brown,B., Brown,M., Bryant,N.P., Buhray,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

```

```

/rpt family="AT_rich"
25572..25592
/rpt family="AT_rich"
26742..26786
/rpt family="TG)n"
26918..26949
/rpt family="AT_rich"
26950..27165
/rpt family="Alu"
30178..30537
/rpt family="L1"
30538..30609
/rpt family="(TAGA)n"
30610..30784
/rpt family="L1"
30785..31082
/rpt family="Alu"
31083..33169
/rpt family="L1"
34810..34951
/rpt family="L1"
35973..37491
/rpt family="L1"
37567..38025
/rpt family="MER2_type"
38012..38078
/rpt family="MER2_type"
38105..38407
/rpt family="Alu"
38745..38870
/rpt family="L2"
38908..39185
/rpt family="Alu"
39432..39497
/rpt family="MIR"
39725..40025
/rpt family="Alu"
43154..43199

Query Match 44.4%; Score 28.4; DB 9; Length 169991;
Best Local Similarity 66.1%; Pred. No. 1.6e+02;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 2 AATTCACCTCCGACCATCCCAAAAGAGAGAGAGTGCATGAAACGAGTAACGTAT 61
Db 129913 AAATCAGTAGCCGAGACTTCACAAAGCTGATGAGTAAGTAATGATCATCATCAATAA 129972

Qy 62 AC 63
Db 129973 AC 129974

RESULT 14
AC026337
LOCUS Homo sapiens 2 BAC RP11-105E24 (Roswell Park Cancer Institute Human
DEFINITION BAC Library) complete sequence.
ACCESSION AC026337
VERSION AC026337.29 GI:14290357
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171004)
Metazoa; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,P.R., Allen,C.,
Alabrooks,S.B., Amarantunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

```

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D.,
Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J.,
Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W.,
Gunaratne, P., Hale, S., Hamilton, K., Kan, J., Harris, C., Harris, K.,
Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,
Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Housi, F.,
Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R.,
Martindale, A., Martinez, E., Massey, S., Mawhinney, E., McLeod, M.P.,
Meador, M., Mei, G., Merscher, S., Metzger, M., Miller, A., Miner, G.,
Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M.,
Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, C., Shim, C.,
Shoostari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A.,
Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A.,
Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B.,
Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D.,
Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R.,
Washington, C., Watlington, S., Williams, G., Williamson, A.,
Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, F., Zhou, J.,
Zorrilla, S., Zuchelapati, R., Weinstein, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 171004)
Worley, K.C.

Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 171004)
Worley, K.C.

Direct Submission
Submitted (05-JUN-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 171004)
Worley, K.C.

Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 5, 2001 this sequence version replaced gi:14277142.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons


```

repeat_region 7264..7294
    /note="5.2 copies 6 mer TATTAA 26% conserved"
repeat_region 7310..7319
    /note="2.5 copies 4 mer GAGG 20% conserved"
repeat_region 7554..7570
    /note="4.5 copies 2 mer GT 25% conserved"
repeat_region 7574..7588
    /note="3.8 copies 4 mer AGGG 23% conserved"
repeat_region 7843..7854
    /note="2.4 copies 5 mer GGGT 24% conserved"
misc_feature 7982..8429
    /note="match: GSS: Em:AQ133669"
repeat_region 8005..8016
    /note="2.4 copies 5 mer AGCT 24% conserved"
misc_feature 8148..8432
    /note="match: GSS: Em:AQ479768"
repeat_region complement(8597..8644)
    /note="MIR repeat: matches 54..111 of consensus"
repeat_region 8645..8802
    /note="MIR repeat: matches 58..212 of consensus"
repeat_region complement(9009..9116)
    /note="MIR3 repeat: matches 48..163 of consensus"
repeat_region 9048..9061
    /note="2.0 copies 7 mer ACTGGG 28% conserved"
repeat_region 9385..9390
    /note="3.3 copies 3 mer TAT 20% conserved"
repeat_region 9665..9777
    /note="12 repeat: matches 2550..2637 of consensus"
repeat_region 9679..9690
    /note="2.0 copies 6 mer CACCTC 24% conserved"
repeat_region 9778..10072
    /note="AluY repeat: matches 1..297 of consensus"
repeat_region 10058..10073
    /note="16.0 copies 1 mer A 32% conserved"
repeat_region 10073..10357
    /note="12 repeat: matches 2637..3036 of consensus"
repeat_region 10152..10167
    /note="5.3 copies 3 mer CTC 23% conserved"
repeat_region 10376..10582
    /note="MER34C repeat: matches 1..233 of consensus"
repeat_region complement(10617..10780)
    /note="Aluub repeat: matches 144..300 of consensus"
repeat_region 10618..10634
    /note="2.8 copies 6 mer TTTTAA 34% conserved"
repeat_region 10698..10709
    /note="2.8 copies 6 mer TTTTAA 34% conserved"

Query Match 44.7%; Score 28.6; DB 9; Length 168860;
Best Local Similarity 72.5%; Pred. No. 14e+02;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0

Qy 1 GAATTCACCTGCCAGACCATCCAAAAAAGAGAGAAAGTGTATGAACA 51
Db 59649 GAATTAACCTCAACCAAGATCACAGGAGAAAGATAGCCATGAACA 59599

```

RESULT	13
AC018678	
LOCUS	169991 bp DNA linear PRI 01-MAR-2002
DEFINITION	Homo sapiens BAC clone RP11-279E17 from 2, complete sequence.
ACCESSION	AC018678
VERSION	AC018678.9 GI:18497216
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Anthropoidea; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	Sulston,J.E. and Waterston,R. 1 (bases 1 to 169991)
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
PUBMED	9847074
REFERENCE	2 (bases 1 to 169991)

CDS
14170..14340,14417..14510,14734..14835,15034..15103,
15187..15286,15377..15471,15562..15641,15764..15842,
15960..16020,16111..16312)
/gene="AT4g03080"
complement(join(11081..11164,11561..11700,11808..11943,
12034..12109,12193..12419,12476..12574,12667..12751,
12834..13072,13152..13322,13483..13684,13779..13950,
14170..14340,14417..14510,14734..14835,15034..15103,
15187..15286,15377..15471,15562..15641,15764..15842,
15960..16020,16111..16312))
/gene="AT4g03080"
/note="contains Pfam Stposphatase signature position 531
to 848 score=376.9, intron 16 is short at 56 nt but no
other splice site can be confidently modeled
strong similarity to similar to
phosphoserine/phosphothreonine phosphatases
contains EST gb:AI994498.1, 144190"
/codon_start=1
/product="putative phospho-ser/thr phosphatase"
/protein_id="CAB77793.1"
/db_xref="GI:7270178"
/db_xref="GOA:O9ZTA4"
/db_xref="SPTREML:O9ZTA4"
/translation="MSKXPLWHPAQYKLETFWDEDDAPGPRCHLUTAVAATKTH
GPKLLFGATALEGGSSVPGIRLAGVINTVHSDYILTRKWRTRKPGEPSPRAAH
AAAVGTMVFOGIGPAGHSTDLLYLDMTNDKFKWRVWVVGDPGPRYGHVMDLV
SQRYLVTDNDKRALSDAWALDTAQKPYVQRLNPDGDRPSARMYAGSARSQDMF
LLCGDRDTLGAFLGILMHNGQWETLAPGAPSPHAAVFGARLHVSGGV
LRGRVTDASVAGNGVDLDTAAGVLDNRGQVTSARGSKGIDQDPSFELMRGRH
GAASVIRIVHGLRGDVLDDPLVAENSTFQSDISSPLLASDRTOQSSSTFRFSYAA
RPPSGSEPSMSEGLSDENSELKEKLEASAEAEVASSVWRAAQLGAGTLDEBFSTS
DASPIVETDGTNEGDRVLRHPRAVVAVAKTVGSLGNVQLSLDQFQNSRRMVP
MNSDVPQPTKTSQKTSQKPGHKKVIAALLPRNWKPPGNKRFGLDYSVEGLCYAA
EQFMHEQTVLQKAPKIVGDLHQFGDLMLRFDYEGFPFGADITYIDYFLGDYV
DRQSHLETITLLALKEIYENHVLIRGNHRAADINALFGFLECIERTKPALYFC
QSGNDGIWATRENQNFYLPALAIENKII CMHGIGISISTVEQIEKIERPITMDA
GSLVMDLLWSDPTDENSELGLRNARGLVTFGDRVTECKNKLOLITRAHECV
MDGPERRAQQLITLSATNYCGTANNAGAILVGRGLVIVPKLHLPLPPILSPENS
PERSGDAMQELNORPPPTTKRFPDPDRSSLAYI"
complement(11081..11164)
/gene="AT4g03080"
/number=1
complement(11165..11560)
/gene="AT4g03080"
/number=1
complement(11561..11700)
/gene="AT4g03080"
/number=2
complement(11701..11807)
/gene="AT4g03080"
/number=2
complement(11808..11943)
/gene="AT4g03080"
/number=3
complement(11944..12033)
/gene="AT4g03080"
/number=3
complement(12034..12109)
/gene="AT4g03080"
/number=4
complement(12110..12192)
/gene="AT4g03080"
/number=4
complement(12193..12419)
/gene="AT4g03080"
/number=5
complement(12420..12475)
/gene="AT4g03080"
/number=5
complement(12476..12574)
/gene="AT4g03080"
/number=6
complement(12575..12666)
/gene="AT4g03080"
complement(12667..12751)
/gene="AT4g03080"
/number=7
complement(12752..12833)
/gene="AT4g03080"
/number=7
complement(12834..13072)
/gene="AT4g03080"
/number=8
complement(13073..13151)
/gene="AT4g03080"
/number=8
complement(13152..13322)
/gene="AT4g03080"
/number=9
complement(13323..13482)
/gene="AT4g03080"
/number=9
complement(13483..13684)
/gene="AT4g03080"
/number=10
complement(13685..13778)
/gene="AT4g03080"
/number=10
complement(13779..13950)
/gene="AT4g03080"
/number=11
complement(13951..14169)
/gene="AT4g03080"
/number=11
complement(14170..14340)
/gene="AT4g03080"
/number=12
complement(14341..14416)
/gene="AT4g03080"
/number=12
complement(14417..14510)
/gene="AT4g03080"
/number=13
complement(14511..14733)
/gene="AT4g03080"
/number=13
complement(14734..14835)
/gene="AT4g03080"
/number=14
complement(14836..15033)
/gene="AT4g03080"
/number=14

Query Match 45.0%; Score 28.8; DB 8; Length 195429;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 16 ACCATGCCAAAAAGAGAGAGAGTGCATGAACACCAAGTAAACGTTATAC 63
|||||
Db 193461 ACCATGACAAAAAGATGCTAAGTGCATGAATACATACGGAATAC 193508
|||||

RESULT 11
AC006842/c

LOCUS
DEFINITION Caenorhabditis elegans clone Y104H12X, *** SEQUENCING IN PROGRESS
AC006842 299015 bp DNA linear HTG 24-FEB-1999
***, 13 unordered pieces.

AC006842 1 GI:4263506
VERSION HTG; HTGS PHASE1.
KEYWORDS Caenorhabditis elegans

SOURCE
ORGANISM Caenorhabditis elegans

REFERENCE
1 (bases 1 to 299015)
AUTHORS Waterston, R.H.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

```

/gene="AT4g03490"
/number=2
complement(11590..12133)
/gene="AT4g03490"
/number=3
complement(12134..12344)
/gene="AT4g03490"
/number=3
complement(12345..12500)
/gene="AT4g03490"
/number=4
complement(12501..12626)
/gene="AT4g03490"
/number=4
complement(12627..12772)
/gene="AT4g03490"
/number=5
complement(12773..12870)
/gene="AT4g03490"
/number=5
complement(12871..13501)
/gene="AT4g03490"
/number=6
misc_feature
13514..15185
/note="marker m1233"
14170..17288
/gene="AT4g03500"
join(14170..14204,15103..15848,15939..16477,16650..17288)
/gene="AT4g03500"
/note="hypothetical protein with some similarity to
ankyrin
similarity to"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77835.1"
/db_xref="GI:7270673"
/db_xref="SPTREMBL:Q9ZT78"
/translation="MNESKISEEISKLGIQSGVSGMRDPVHLDRRVAENHQAIIPR
NNVRFPSLDLSTLDETSETPKMDPTMAAVRAGKENVLRNNSYISVAPTLVNDRG
NTILHLAASGHVSLVRYIIQKCPGLLKNMGEVALHLAAEAGHLVDVWMLIDFIN
DISCTILPVAKRIFYAKNNQDPTALHVKRGHEVAVSYLVSAAXLSFVANKDGFSP
LYLAIAGHTSLVTTWCHGTNELSKVGGRSIVHAALKANRKOIIDLALLSKDASLINL
RDSGRSLSGASIGYVQGSYLFQKNDKRVYSDGGLPFTHMAKYGHVILSEIL
KHCPAEILLDRGQNILHLAAYGKLVKIFILSCCKNNKKLINEQDVNGVPTLH
LATINHPKVVSMFTDHRVLDLKRNYIGFTALDVABENIDSSYIVHQELTWMALINA
GAPKSTPIENRSPKPKDGGYKDRVNTLMVATVATMTFTAGFTLPGGYNDSFP
HLGMVIAKRTAFQVFLVCDTLAMYGSIIITIVALLWAQGLDLSILKAFNIALPLGL
ALTMGSIAMAGTYAVSHPLILGYFVLGIGIIFLIVLLLLVPVSPYAHQAQLLRH
IFYYPVFLKLLAAGDNKNIQDVVYASDE"
14170..14204
/gene="AT4g03500"
/number=1
14205..15102
/gene="AT4g03500"
/number=1
15103..15848
exon
Query Match 45.0%; Score 28.8; DB 8; Length 159629;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 16 ACCATGCCAAAAGAGAGAGAGGTCATGAACACCAAGTAACGTTATAC 63
|||||
Db 3032 ACCATGACAAAAGAAATGCTAAGGTCATGAATACATAGCGGAATAC 3079
|||||
RESULT 10
ATCHRIV8
LOCUS 195429 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 8.
ACCESSION AL161496
VERSION AL161496.2 GI:7270176
KEYWORDS

```

SOURCE ORGANISM

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE AUTHORS

1 (bases 1 to 195429)
Spiegel, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M.,
Vil, D.M., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Sekher, M., Schutz, K., See, L.H., Swaby, I.,
Habermann, K., Dedhia, N.N., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 195429)

AUTHORS

EU Arabidopsis sequencing project.

TITLE

Direct Submission

JOURNAL

Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

COMMENT

Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV7 at the 5' end and an
overlap with ATCHRIV9 at the 3' end.

FEATURES

Location/Qualifiers

1..195429
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
9588..10766
/gene="AT4g03070"
join(9588..9967,10106..10445,10518..10766)
/gene="AT4g03070"
/note="most similar to T419.6 and T419.7, contains Pfam
iron/ascorbate family of oxidoreductases signature
position 26 to 287 score=22.4
similarity to
contains EST gb:AI98452.1, AI999675.1"
/codon_start=1
/product="putative oxidoreductase"
/protein_id="CAB77792.1"
/db_xref="GI:7270177"
/db_xref="GOA:Q9ZTA3"
/translation="MDSDFVPSVFCPLVIDSDNLKPGSKWDEVTADVLKALD
YGCFAEAFDKLSVELNRSVFEMEDLELPIPTKQNVSSKPFHYLCHLYESLGIN
DANVIEKVNDFTOQLWPDHGNKISLSETHLFSQVLELDLWVRMINMESFGIENYIDE
HLNSTYLTLMKYTSPDDDDDEETKGLRSHTDKNIITLHQVQDGLVETKID
DKWIKVPSQDSVLVMVYDGLCALLNGRLSHPYRVTMGKTRYSGLPSIPKTVI
IDSPLELVDKHEPRIFKPFYEYDFLHPFQTEAGRIQAQSALHFAAF"
9588..9967
/gene="AT4g03070"
/number=1
9968..10105
/gene="AT4g03070"
/number=1
10106..10445
/gene="AT4g03070"
/number=2
10446..10517
/gene="AT4g03070"
/number=2
10518..10766
/gene="AT4g03070"
/number=3
11081..16312
/gene="AT4g03080"
complement(join(11081..11164,11561..11700,11808..11943,
12034..12109,12193..12419,12476..12574,12667..12751,
12834..13072,13152..13322,13483..13684,13779..13950,

45.0%; Score 28.8; DB 8; Length 95519;
ity 75.0%; Pred. No. 1.3e+02;
servative 0; Mismatches 13; Indels 0; Gaps 0;
CAAAAAGAGAGAGAGGTCATGAAACCAAGTAAAGTTATAC 63
|||||
CAAAAAGAGAGGTCATGAAACCAAGTAAAGTTATAC 49982

QY 1 GAATTCACCTCCAGACCATCCAAAAGAGAGAGAGGCTCATGAAACCAAGTAAAGCTTA 60
 DB 12475 GAATTTTCCTGTGACGCGGTGGTTAAAGAGAGAGAGAGAGCTGAAATACAGAAATGTTA 12534

QY 61 TACG 64
 DB 12535 AAAG 12538

RESULT 7
 LOCUS AF036699/c 43764 bp DNA linear INV 16-JUN-2003
 DEFINITION Caenorhabditis elegans cosmid F58F6, complete sequence.
 ACCESSION AF036699
 VERSION AF036699.2 GI:4895094
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 43764)
 Wilson, R.
 Genomic sequencing of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99069613
 9851916
 2 (bases 1 to 43764)
 Dante, M., Kramer, J. and Gibson, A.
 The sequence of C. elegans cosmid F58F6
 Unpublished (2001)
 3 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (03-DEC-1997) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 4 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (27-MAY-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (05-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (22-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (28-AUG-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (19-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 43764)
 Waterston, R.
 Direct Submission
 Submitted (07-APR-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 10 (bases 1 to 43764)

Wilson, R.
 Direct Submission
 Submitted (16-JUN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 On May 27, 1999 this sequence version replaced gi:2662597.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1EQ, England
 email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=F58F6;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F53H1, 3000 bp overlap; the 3' cosmid is C07B3,
 200 bp overlap. Actual start of this cosmid is at base position 1
 of F58F6; actual end is at 43764 of F58F6.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohasa
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://worfdb.dicf.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers
 1. .43764
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="IV"
 /clone="F58F6"
 4765. .6410
 /gene="F58F6.6"
 join(4765..4797,5041..5184,5298..5418,5515..5644,
 5695..5878,5924..6076,6128..6211,6264..6410)
 /gene="F58F6.6"
 /standard_name="F58F6.6"
 /note="contains similarity to Interpro domain IPR006874
 (protein of unknown function DUF621)"
 /codon_start=1
 /product="Hypothetical protein F58F6.6"
 /protein_id="AA988362.1"
 /db_xref="GI:2662603"
 /db_xref="WormBase:F58F6.6"
 /translation="MEKRLGPMRELLDSDSEKNWLYYSVVSFFVVTIGATILTM"

FEATURES
 source
 gene
 CDS

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this clone sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=ZC196;class=Sequence

NEIGHBORING CLONE INFORMATION

The 5' clone is ZC178, 200 bp overlap; the 3' clone is B0507, 200 bp overlap. Actual start of this clone is at base position 1 of ZC196; actual end is at 3481 of B0507.

NOTES:

Coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1123, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. TRNAS are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers

```

FEATURES             source
    source
1. .33010
   /organism="Caenorhabditis elegans"
   /mol_type="genomic DNA"
   /strain="Bristol N2"
   /db_xref="taxon:6239"
   /chromosome="V"
   /clone="ZC196"
   /size="3767"
   /gene="glr-5"
   /genes="glr-5"
   join(3767..3848,4081..4230,4284..4375,4420..4579,
4766..5490,5770..6292,6338..6451,6521..6649,6693..7095,
7159..7252,7330..7558,7611..7718)
   /genes="glr-5"
   /standard_name="ZC196.7"
   /note="contains similarity to Pfam domains PF01094
(Receptor family ligand binding region), PF00060
(Ligand-gated ionic channels); coded for by the following
C. elegans cDNAs: AF318609, Yk87h12.5, Yk1222f9.5,
Yk1302h05.3, Yk1302h05.5, Yk1337e03.3, Yk1337e03.5"
   /codon_start=1
   /product="Glutamate receptor family (ampa) protein 5"
   /protein_id="AAB52291.1"
   /db_xref="GI:1938466"
   /db_xref="WormBase:ZC196.7"
   /translation="MQFNYYIFSIITLHVFTSCYAKIGAIFFRERDAHIEAIVR
YSVWLSQRTGIDIVDIYIVDYLDAIKKAKCKLEKHIVALLGGSHAALNAQL
ERITDDIDPFLAIDDLDRDMGSKIDFWPQLFEAVDMFHWNRWIRVLYVEGD
ERIRRLQOLLESRYASIRFYLKVHGDGYMKAAROVKELECRLLHKKDCSEFSL
VDMNPEHTYFLASLQMLGLKHWPLLNMLSTMDVLFRLYNHARTFISYPVDST
FLTENRDAFNSHPEKHTKMAWKTDSNKLKMEAVTFDVAFAFANVFNELSSH
QMNDVPTQLCRSRNRKQYHGKSLDINVDLGLSGDLRLNGHPLKSNFSKRI
MLKKNHYLRDNDIGFWEFATNVHNMDSGSAQLQNVQVSDLEKPHFRVITINERIV
MLKKNHYLRDANSFEFCIDLLAELSKGLGFTYTHAVXDKYKNDKYGNSGDGM
GEILRGAENAVAPLTVNYRSEAVDFTKPFLSIGISILYKVPDQDQDPLFSLNPLS
WQIWTATATSIITVLGMYFVANVTPYEWNLNFSCTAHEPHFAAFATNQEAIPVNS

```

gene

CDS

```

/join(13756..13923,13866..14206,14251..14412,14539..14625,
14788..14955,15004..15051,15096..15215,15259..15336,
15378..15443,15490..15588,15712..15826,15873..15996)
/gene="ZC196.6"
/standard_name="ZC196.6"
/note="contains similarity to Mycoplasma genitalium
Cytochrome high molecular weight protein 1 (Cytochrome
accessory protein 1); SW:HMW1 MYCGE; coded for by the
following C. elegans cDNAs: OSTF185C5_1, OSTF185C5_1,
OSTF185C5_2"
/codon_start=1
/product="Hypothetical protein ZC196.6"
/protein_id="AAB52297.2"
/db_xref="GI:38176081"
/db_xref="WormBase:ZC196.6"
/translation="MSTFKKCHPPLSNXSGEMLKIROGTHIRSEPVQFSNQPPVNP
PNNVQFRFTDGLPKHKNRPVWPNSSVFRCTVTVYVNLKLPQVQKQKFKESLP
VMNPKSPREDPVVPEIKTSSGKGVQHVQSKLHSENEGVPDPLADVDVHRP
FVNTQVGLNSGTEKPDVTAKNLEPSKITVNEETKPDQDQIQIISHLDDAKEDVD
SVLELNDNQCIDDINKTSEVKEIKFEETVAVPDNEQFVNSFNVDNCEKE
TSNEDQONQDVVAETMEKPGNLTRGFTENRLLIEACLDGIEHSNIVAIETREDE

```

gene

CDS

```

/join(13756..13923,13866..14206,14251..14412,14539..14625,
14788..14955,15004..15051,15096..15215,15259..15336,
15378..15443,15490..15588,15712..15826,15873..15996)
/gene="ZC196.6"
/standard_name="ZC196.6"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621))"
/codon_start=1
/product="Hypothetical protein ZC196.9"
/protein_id="AAB52299.1"
/db_xref="GI:1938474"
/db_xref="WormBase:ZC196.9"
/translation="MKTDEALHMVLSHENVLSLNFNGVTGKFLNPKQYIISFKFP
LGNLKNISMLILDIDSDSEKNLYSVSVLTVSTIGATILMSFLVLSLILWHRFKS
MRFFLQUTTSAPFVISAANLFINIPALSLFSEKVTOSRMFVYVILDFCHYSIL
FNSLVIAQRAFVFFLRNLDRFRFSKYIYINISSYIYAFGEYVAFMNSNCOYRGE
LARKYQLICETINAYNVITMTTPTVQLMTLMQFLPLSLIDIIYAIATVRIYKNG
SLTKXNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILVFKRFTVNTMILAGAA
TPSPFPFTSKIRKLVSTRVSAVSSQSPNSIQVRAARTLE"
/size="13756"
/gene="ZC196.9"
/complement(join(10750..10896,10949..11032,11084..11236,
11282..11465,11514..11643,11740..11860,11974..12117,
12406..12513,12751..12785,13111..13129))
/gene="ZC196.9"
/standard_name="ZC196.9"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621))"
/codon_start=1
/product="Hypothetical protein ZC196.9"
/protein_id="AAB52299.1"
/db_xref="GI:1938474"
/db_xref="WormBase:ZC196.9"
/translation="MKTDEALHMVLSHENVLSLNFNGVTGKFLNPKQYIISFKFP
LGNLKNISMLILDIDSDSEKNLYSVSVLTVSTIGATILMSFLVLSLILWHRFKS
MRFFLQUTTSAPFVISAANLFINIPALSLFSEKVTOSRMFVYVILDFCHYSIL
FNSLVIAQRAFVFFLRNLDRFRFSKYIYINISSYIYAFGEYVAFMNSNCOYRGE
LARKYQLICETINAYNVITMTTPTVQLMTLMQFLPLSLIDIIYAIATVRIYKNG
SLTKXNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILVFKRFTVNTMILAGAA
TPSPFPFTSKIRKLVSTRVSAVSSQSPNSIQVRAARTLE"
/size="13756"
/gene="ZC196.9"
/complement(join(10750..10896,10949..11032,11084..11236,
11282..11465,11514..11643,11740..11860,11974..12117,
12406..12513,12751..12785,13111..13129))
/gene="ZC196.9"
/standard_name="ZC196.9"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621))"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following
C. elegans cDNAs: OSTF164G3_1, OSTF164G3_1"
/codon_start=1
/product="Hypothetical protein ZC196.8"
/protein_id="AAB52298.2"
/db_xref="GI:38176082"
/db_xref="WormBase:ZC196.8"
/translation="MRFFWFLTQUTTSAPFVISAANLFINIPALSLFSEKVTOSQLFY
PFSYLIDFCHYSILFNSLVISIQRAFVFFLRNTWQFFESKAIYIWLASIIYFARGVE
YALMSSNCQGYDKTSKNYQLVCEADNRGIMKTVAIPAGIQFVETLHIVIFLILII
YAIATRIIYMGNTLSKNETLILKQAFVFLIFQASSCVFLFAQSQVITNVGAFILV
RFVNTMILAGAAATPSFPFTSKIRKLVSTRVSAASLSGLSPNSIQVRAARTLE"
/complement(join(10750..11129))
/gene="ZC196.8"
/complement(join(8426..8572,8623..8706,8759..8911,
8961..9141,9192..9321,9368..9488))
/gene="ZC196.8"
/standard_name="ZC196.8"
/note="contains similarity to Pfam domain PF04789 (Protein
of unknown function (DUF621)); coded for by the following

```

```

Query Match      45.9%; Score 29.4; DB 8; Length 88813;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2 AATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCACTAAGTTAT 61
    |||||
Db 22756 AATTCATCTGCAAGACTCCAGAAAAATAGATGAGCAGGTAATGCAACCAAACTGCTAC 22815

QY 62 ACG 64
Db 22816 ATG 22818

RESULT 5
AC016779 140234 bp DNA linear PLN 19-APR-2003
LOCUS Genomic sequence for Oryza sativa clone 10N6, complete sequence.
DEFINITION AC016779
ACCESSION AC016779.5 GI:30025033
VERSION HTG.
KEYWORDS Oryza sativa
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaeae; Oryza.
REFERENCE 1 (bases 1 to 140234)
AUTHORS McCombie, W.R., de la Bastide, M., Spiegel, L., Huang, E., Preston, R.,
Nascimento, L., Zutavern, T., Balija, V., Bell, M., Miller, B.,
Muller, S., Katzenberger, F., Yang, C., Dike, S., O'Shaughnessy, A.,
Dedhia, N. and Palmer, L.
GENOMIC SEQUENCE FOR Oryza sativa, clone 10N6, complete sequence
Unpublished
REFERENCE 2 (bases 1 to 140234)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2000) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
REFERENCE 3 (bases 1 to 140234)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2003) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
COMMENT On Apr 19, 2003 this sequence version replaced gi:1094443.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
FEATURES
    source
        1..140234
            /organism="Oryza sativa"
            /mol_type="genomic DNA"
            /strain="Teqing"
            /db_xref="taxon:4530"
            /chromosome="5"
            /clone="10N6"
            2541..2622
            /note="The sequence is a dinucleotide (TA) repeat in which
            the exact number of repeat units is unknown."
            complement (82075..82235)
            /note="The region is covered by a single subclone which
            was sequenced with dye terminator chemistry. The quality
            is below phred30 and there are ambiguous bases at
            positions 82079 and 82114."
            positions 82079 and 82114."
            83100..83530
            /note="The sequence is covered by a PCR product which was
            amplified by a high fidelity polymerase."

misc_feature
misc_feature
misc_feature
ORIGIN

```

```

Query Match      45.9%; Score 29.4; DB 8; Length 140234;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 42; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2 AATTCACCTGCCAGACCATGCCAAAAAGAGAGAAAGGTGATGAACCACTAAGTTAT 61
    |||||
Db 79808 AATTCATCTGCAAGACTCCAGAAAAATAGATGAGCAGGTAATGCAACCAAACTGCTAC 79867

QY 62 ACG 64
Db 79868 ATG 79870

RESULT 6
U97007 33010 bp DNA linear INV 03-DSC-2003
LOCUS Caenorhabditis elegans cosmid ZC196, complete sequence.
DEFINITION U97007
ACCESSION U97007
VERSION U97007.1 GI:1938465
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 33010)
AUTHORS Wilson, R.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
JOURNAL MEDLINE
PUBMED 9851916
REFERENCE 2 (bases 1 to 33010)
AUTHORS Murray, J.
TITLE The sequence of C. elegans cosmid ZC196
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 33010)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1997) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 4 (bases 1 to 33010)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 5 (bases 1 to 33010)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 6 (bases 1 to 33010)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2003) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 7 (bases 1 to 33010)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (03-DSC-2003) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

```

```
JOURNAL Patent: WO 0222834-A 13 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..6148
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Construct Sequence"
ORIGIN
Query Match 100.0%; Score 64; DB 6; Length 6148;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 60
Db 1 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 60
QY 61 TAGC 64
Db 61 TAGC 64
RESULT 2
LOCUS AX468461 66 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 4 from Patent WO0222834.
ACCESSION AX468461
VERSION AX468461.1 GI:21901297
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE
  1 Siebel, C. and Brennan, T.J.
  Methods of producing cells and animals comprising targeted gene
  modifications
  JOURNAL Patent: WO 0222834-A 4 21-MAR-2002;
  Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..66
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Primer"
ORIGIN
Query Match 100.0%; Score 64; DB 6; Length 66;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 60
Db 3 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 62
QY 61 TAGC 64
Db 63 TAGC 66
RESULT 3
LOCUS AX468470 6148 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 13 from Patent WO0222834.
ACCESSION AX468470
VERSION AX468470.1 GI:21901306
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE
  1 Siebel, C. and Brennan, T.J.
  Methods of producing cells and animals comprising targeted gene
  modifications
JOURNAL Patent: WO 0222834-A 13 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..6148
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Construct Sequence"
ORIGIN
Query Match 100.0%; Score 64; DB 6; Length 6148;
Best Local Similarity 100.0%; Pred. No. 2.4e-09;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 60
Db 1 GAATTCACCTGCCAGACCATGCCAAAAGAGAGAGAAAGGTTCATGAACCAAGTAAACGTTA 60
QY 61 TAGC 64
Db 61 TAGC 64
RESULT 4
LOCUS AC105772 88813 bp DNA linear PLN 09-AUG-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 5 clone
OJ1504_G04, complete sequence.
ACCESSION AC105772
VERSION AC105772.2 GI:33563204
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE
  1 (bases 1 to 88813)
  Chow, T.-Y., Hsing, Y.-I. C., Chen, C.-S., Chen, H.-H., Liu, S.-M.,
  Chao, Y.-T., Chang, S.-J., Chen, H.-C., Chen, S.-K., Chen, T.-R.,
  Chen, Y.-L., Cheng, C.-H., Chung, C.-I., Han, S.-Y., Hsiao, S.-H.,
  Hsiung, J.-N., Hsu, C.-H., Huang, J.-J., Kau, P.-I., Lee, M.-C.,
  Lee, H.-L., Li, Y.-F., Lin, S.-J., Lin, Y.-C., Wu, S.-W., Yu, C.-Y.,
  Yu, S.-W., Wu, H.-P. and Shaw, J.-F.
  Oryza sativa BAC OJ1504_G04 genomic sequence
  JOURNAL Unpublished
  REFERENCE
  2 (bases 1 to 88813)
  Chow, T.-Y. and Hsing, Y.-I. C.
  Direct Submission
  TITLE Submitted (10-JAN-2002) Institute of Botany, Academia Sinica, 128,
  JOURNAL Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
  REFERENCE
  3 (bases 1 to 88813)
  Hsing, Y.-I. C. and Chow, T.-Y.
  Direct Submission
  TITLE Submitted (28-MAR-2003) Institute of Botany, Academia Sinica, 128,
  JOURNAL Section 2, Yien-chu-Yuan Road, Nankang, Taipei 11529, Taiwan
  REFERENCE
  4 (bases 1 to 88813)
  Chow, T.-Y. and Hsing, Y.-I. C.
  Direct Submission
  TITLE Submitted (09-AUG-2003) Institute of Botany, Academia Sinica, 128,
  JOURNAL Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
  COMMENT On Aug 9, 2003 this sequence version replaced gi:18104691.
  The nucleotide sequence of this BAC clone was generated by
  combining Monsanto and ASPGC-Taiwan sequencing data.
FEATURES
  source
    Location/Qualifiers
      1..88813
      /organism="Oryza sativa (japonica cultivar-group)"
      /mol_type="genomic DNA"
      /cultivar="Nipponbare"
      /db_xref="taxon:39947"
      /chromosome="5"
      /clone="OJ1504_G04"
ORIGIN
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 22, 2004, 10:25:20 ; Search time 357.763 Seconds
(without alignments)
7753.612 Million cell updates/sec

Title: US-09-954-483B-3
Perfect score: 64
Sequence: 1 gaattcaactgcagaccat.....gaaaccagtaacgttatcacg 64

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_ph :
7: gb_pl :
8: gb_pr :
9: gb_ro :
10: gb_sy :
11: gb_sy :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_un :
30: em_vt :
31: em_vt :
32: em_vt :
33: em_vt :
34: em_vt :
35: em_vt :
36: em_vt :
37: em_vt :
38: em_vt :
39: em_vt :
40: em_vt :
41: em_vt :

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	64	100.0	64	6	AX468460	Sequence
2	64	100.0	66	6	AX468461	Sequence
3	64	100.0	66	6	AX468470	Sequence
4	29.4	45.9	88813	8	AC105772	Oryza sat
5	29.4	45.9	140234	8	AC016779	Genomic s
6	28.8	45.0	33010	3	U97007	Caenorhabdi
7	28.8	45.0	43764	3	AF036699	Caenorhab
8	28.8	45.0	95519	8	AF071527	Arabidops
9	28.8	45.0	159629	8	ATCHRIV9	Arabidops
10	28.8	45.0	195429	8	ATCHRIV8	Arabidops
11	28.8	45.0	229015	2	AC006842	Caenorhab
12	28.6	44.7	168860	9	AL160175	Human DNA
13	28.4	44.4	169991	9	AC018678	Homo sapi
14	28.4	44.4	171004	9	AC026337	Homo sapi
15	28.4	44.4	187547	9	AC078958	Homo sapi
16	27.8	43.4	180500	2	AY178789	Canis fam
17	27.8	43.4	275089	2	AC128134	Rattus no
18	27.6	43.1	36625	3	CEFS6H6	Caenorhabdi
19	27.6	43.1	174055	9	AC104358	Homo sapi
20	27.6	43.1	200738	9	AF005622	Homo sapi
21	27.6	43.1	231236	2	AC096369	Rattus no
22	27.6	43.1	243489	2	AC096036	Rattus no
23	27.4	42.8	3319	8	AY090988	Arabidops
24	27.4	42.8	48198	2	AC008801	Homo sapi
25	27.4	42.8	100424	8	AC005700	Arabidops
26	27.4	42.8	110000	1	AP006628	Continuation (4 of
27	27.4	42.8	110000	1	AP006628	Continuation (5 of
28	27.4	42.8	131398	9	HS445C9	Human DNA s
29	27.4	42.8	184065	2	AC108371	Pan trogl
30	27.2	42.5	3210	10	BC028971	Mus muscu
31	27.2	42.5	20792	9	CNS07EFE	Human chr
32	27.2	42.5	135810	9	AC023992	Homo sapi
33	27.2	42.5	137176	9	AC005017	Homo sapi
34	27.2	42.5	165384	2	AC055735	Homo sapi
35	27.2	42.5	170892	2	AC068438	Homo sapi
36	27.2	42.5	175120	9	AC004687	Homo sapi
37	27.2	42.5	178419	2	EX005174	Danio rer
38	27.2	42.5	196074	2	AC122218	Mus muscu
39	27.2	42.5	196703	2	AC096851	Pan trogl
40	27.2	42.5	200747	2	AC107681	Mus muscu
41	27	42.2	538	11	G92838	S203P6142RB
42	27	42.2	62466	9	AC093785	Homo sapi
43	26.8	41.9	110000	2	EX119990	Continuation (3 of
44	26.8	41.9	123385	1	EX640452	Bordetell
45	26.8	41.9	123668	9	AL390728	Human DNA

ALIGNMENTS

RESULT 1
AX468460
LOCUS AX468460
DEFINITION Sequence 3 from Patent WO022834.
ACCESSION AX468460
VERSION AX468460.1 GI:21901296
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Siebel, C. and Brennan, T.J.
TITLE Methods of producing cells and animals comprising targeted gene
modifications
JOURNAL Patent: WO 022834-A 3 21-MAR-2002;
64 bp DNA linear PAT 16-JUL-2002

ORIGIN

Query Match 40.0%; Score 47.6; DB 13; Length 354;
Best Local Similarity 72.1%; Pred. No. 7.1e-05;
Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CTCACAAATCCCGGCATTTCTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGC 78
DB 8 CTCGGAGGCGCGCATTTCTGCACGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGC 67

QY 79 GCTCACAATTCGCGGCGCTTTGACCT 104
DB 68 TTCTCATCTCCGGCGCTTTGACCT 93

Search completed: April 22, 2004, 12:56:03
Job time : 732.893 secs

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Tagi, K., Iomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Buit, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustinch, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konasaya, A.,
Kurukchin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., Mckenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perteau, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, F., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L.G., Wnshaw-Boris, A., Yanagisawa, M., Yang, I.,
Wells, C., Wilming, L.G., Zolman, M., Zhu, Y., Zimmer, A., Carninci, P.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kigawa, I., Miyazaki, A., Sakai, K., Sasaki, P., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

MEDLINE

22354683

PUBMED

12468851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ono, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Takashi Ishikawa (Department of Surgery
2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama
236-0004 Japan) whose assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
Location/Qualifiers
i. .354
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

FEATURES

source

assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

```

source
  1. 353
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="I830045M08"
  /tissue_type="bone marrow"
  /cell_type="macrophage"
  /clone_lib="RIKEN full-length enriched, bone marrow
  macrophage"

ORIGIN
  Query Match 40.0%; Score 47.6; DB 13; Length 353;
  Best Local Similarity 72.1%; Pred. No. 7.1e-05;
  Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CTCACATCCCGGCATTCGCGAGCTTCAAAGCGCAGCTGCGCGCGCTATTGTGAGC 78
    |||
Db 45 CTCGAGGCGCGGCATTCGCGAGCTTCAAAGCGCAGCTGCGCGCGCTATTGTCTCC 104
    |||

QY 79 GCTCACATTCGCGGCTTCGACCT 104
    |||
Db 105 TTCTCATCTCCGGCGCTTCGACCT 130
    |||

```

RESULT 14

BY068766 RIKEN full-length enriched, 17 days embryo heart Mus EST 06-DEC-2002
 musculus cDNA clone I920061P15 5', mRNA sequence.

BY068766 GI:26171926

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 354);
 Okazaki, I., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaado, I., Osato, N., Saito, R., Suzuki, H., Yamana, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Faven, W. J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Sempile, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp/
 URL: <http://genome.gsc.riken.go.jp/>
 Aizawa, K., Akimura, T., Atakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES

source

```

Location/Qualifiers
  1..354
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="I920061P15"
  /tissue_type="heart"
  /dev_stage="17 days embryo"
  /clone_lib="RIKEN full-length enriched, 17 days embryo
  heart"

```

ORIGIN

```

Query Match 40.0%; Score 47.6; DB 13; Length 354;
Best Local Similarity 72.1%; Pred. No. 7.1e-05;
Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CTCACATCCCGGCATTCGCGAGCTTCAAAGCGCAGCTGCGCGCGCTATTGTGAGC 78
    |||
Db 34 CTCGAGGCGCGGCATTCGCGAGCTTCAAAGCGCAGCTGCGCGCGCTATTGTCTCC 93
    |||

QY 79 GCTCACATTCGCGGCTTCGACCT 104
    |||
Db 94 TTCTCATCTCCGGCGCTTCGACCT 119
    |||

```

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

BY308363 RIKEN full-length enriched, stroma cell Mus musculus cDNA
 clone I320005B10 5', mRNA sequence.
 BY308363.1 GI:26498700
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 354)

REFERENCE
AUTHORS

1 (bases 1 to 346)
 Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Adachi,J., Bono,H., Kondo,S.,
 Nikaado,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gliss,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sautana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
 Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
 Rogers,J., Birney,E. and Hayashizaki,Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 12466851
 22354683
 12466851

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
 Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
 Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,K.,
 Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
 Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
 Hayashizaki,Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES
source

Location/Qualifiers

1. 346
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L630086K24"

/sex="male"

/tissue type="brain"

/dev stage="adult"

/lab host="DH10B"

/clone lib="RIKEN full-length enriched, adult male brain"

/note="Site 1: Sali; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTAN 3']. cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGATCTCGAGTTAATAATAATCCCCCCCCCCC 3']. cDNA
 was cloned into the XhoI and BamHI sites. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I"

ORIGIN

Query Match 40.0%; Score 47.6; DB 13; Length 346;
 Best Local Similarity 72.1%; Pred. No. 7.1e-05;
 Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
 QY 19 CTCACATCCGGCATCTCGACGCTTCAAAAGCGACGCTGCGCGGTATTGTGAGC 78
 DB 9 CTCGGAGGCGCGCATCTTCGACGCTTCAAAAGCGACGCTGCGCGGTATTGTCTC 68
 QY 79 GCTCACATTCGGGGCGCTTTTCGACCT 104
 DB 69 TTCTCATCTCCGGCGCTTTTCGACCT 94

RESULT 12

BY085570

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY085570 RIKEN full-length enriched, 14 days embryo whole body Mus
 musculus cDNA clone K630053H15 5', mRNA sequence.
 BY085570
 BY085570.1 GI:26195478
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 350)
 Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaado,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
 Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
 Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
 Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
 Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
 Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
 Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
 Gariboldi,M., Gliss,C., Godzik,A., Gough,J., Grimmond,S.,
 Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
 Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
 Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
 Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
 Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
 Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
 Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
 Sautana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
 Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
 Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
 Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
 Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
 Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,

Query Match 40.0%; Score 47.6; DB 13; Length 310; Best Local Similarity 72.1%; Pred. No. 6.8e-05; Mismatches 24; Indels 0; Gaps 0; Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

19 CTCACATCCCGGCGATTCGCAAGCTTCACAAAGCGCAGCTCTCGCGCGCTATTGTGAGC 78
 9 CTCGGAGGCGCGGCGATTCGCAAGCTTCACAAAGCGCAGCTCTCGCGCGCTATTGTCTCTC 68

QY 79 GCTCACAATCCGGCGCTTCGACCT 104
 Db 69 TTCCTCATCTCCGGCGCTTCGACCT 94

RESULT 10
 BY124531
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojoberi, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Baisel, K. W., Blake, J. A., Bratt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Jarvis, E. D., Kanai, A., Kustanich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Q, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

22354683
 12466851

CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Itoh, M., Kawai, J., Konno, H., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
 1. .326
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L630026B14"
 /sex="male"
 /tissue_type="brain"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, adult male brain"
 /note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'
 GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTT 3'. cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'
 GAGAGAGATCTCGATTAATTAATATCCCGCCCCCCC 3'. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1"

ORIGIN

Query Match 40.0%; Score 47.6; DB 13; Length 326;
 Best Local Similarity 72.1%; Pred. No. 6.9e-05;
 Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

19 CTCACATCCCGGCGATTCGCAAGCTTCACAAAGCGCAGCTCTCGCGCGCTATTGTGAGC 78
 26 CTCGGAGGCGCGGCGATTCGCAAGCTTCACAAAGCGCAGCTCTCGCGCGCTATTGTCTCTC 85
 QY 79 GCTCACAATCCGGCGCTTCGACCT 104
 Db 86 TTCCTCATCTCCGGCGCTTCGACCT 111

RESULT 11

BY130145
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BY130145 346 bp mRNA linear EST 06-DEC-2002
 cDNA clone L630086K24 5', mRNA sequence.
 BY130145
 BY130145.1 GI:26241246
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosewa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers
1. .267
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K530325N08"
/tissue type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

Query Match 40.0%; Score 47.6; DB 13; Length 267;
Best Local Similarity 72.1%; Pred. No. 6.4e-05;
Matches 62; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CTCACAATCCGGGCGGCTTCGCAAGCTTCAAAAGCGCAGCTGCGCGGCTATTGTGAGC 78
Db 52 CTCGGAGGCGCGGCGATTCGACGCTTCARAGGCGACGCTGCGGCGCTTCTCCTC 111
QY 79 GTCACAATCCGGGCGGCTTCGACCT 104
Db 112 TTCTCATCTCCGGGCGCTTCGACCT 137

RESULT 9
BY066197
LOCUS
DEFINITION
amion Mus musculus cDNA clone 1920047124 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 310)
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Eult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,

Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.P., Forrest, A., Frazer, K.S., Gaasterland, T.,
Garioldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Santalin, A., Schneider, C., Sempie, C.A., Setou, M., Shimada, K.,
Sundell, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, K.,
Vardaro, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayashizaki, Y., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kigawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
23354683
12468851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
Location/Qualifiers
1. .310
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I920047124"
/sex="female"
/tissue type="amion"
/dev_stage="17 days pregnant adult"
/clone_lib="RIKEN full-length enriched, 17 days pregnant
adult female amion"

FEATURES
SOURCE

ORIGIN

CONTACT: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

FEATURES
SOURCE

```

Db 1 GGGCTCCGAGGCCCGGCATCTCGACGTTCAAAAGCGCACGTCCTGCCGGCTGTCTC 60
Qy 76 AGCGCTCACAAATCCCGGGCCTTTTCGACCT 104
Db 61 CTCCTCTCATCTCCGGGCTTTTCGACCT 89

RESULT 4
BUS25591
LOCUS
DEFINITION
AGENCOURT 10120568 NIH_MGC_144 Mus musculus cDNA clone
IMAGE:6534000 5', mRNA sequence.
ACCESSION
BUS25591
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1023)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2691 row: g column: 24
High quality sequence stop: 638.
Location/Qualifiers
1..1023
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6534000"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_144"
/notes="Organ: Brain; Vector: pDNR-LIB; Site:1: SfiI
(ggccattggcc); Site:2: SfiI (ggcgctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGCGCGGCATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.2-0.5
Kb size fraction (other fractions present in NIH_MGC_143).
Library created in the laboratory of M. Brownstein (NIMH,
NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 40.5%; Score 48.2; DB 13; Length 1023;
Best Local Similarity 89.9%; Pred. No. 6.8e-05;
Matches 65; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 12 GTGAGCGCTCACAAATCCCGGCATCTCGCAAGCTTCAAAAGCGCACGTCCTGCCGGCTAT 71
Db 1 GGGGGTCTCTCGAGGCCCGGCATCTTGACGCTTCAAAAGCGCACGTCCTGCCGGCTGT 60
Qy 72 TGTGAGCGCTCACAAATTCGGGCTTTTCGACCT 104
Db 61 TCTCTCTTCTCATCTCCGGGCTTTTCGACCT 93

RESULT 5
BUS71341
LOCUS
DEFINITION
AGENCOURT 10372555 NIH_MGC_144 Mus musculus cDNA clone
IMAGE:6623467 5', mRNA sequence.

REFERENCE
1 (bases 1 to 752)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.

```

```

ACCESSION
BUS71341
VERSION
BUS71341.1 GI:22921641
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 752)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2878 row: g column: 19
High quality sequence stop: 465.
Location/Qualifiers
1..752
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6623467"
/lab_host="DH10B (TI-phage-resistant)"
/clone_lib="NIH_MGC_144"
/notes="Organ: Brain; Vector: pDNR-LIB; Site:1: SfiI
(ggccattggcc); Site:2: SfiI (ggcgctcgcc); cDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGCGCGGCATG-dt(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.2-0.5
Kb size fraction (other fractions present in NIH_MGC_143).
Library created in the laboratory of M. Brownstein (NIMH,
NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 40.2%; Score 47.8; DB 13; Length 752;
Best Local Similarity 70.3%; Pred. No. 8.1e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 14 GAGCGCTCACAAATCCCGGCATCTTCGCAAGCTTCAAAAGCGCACGTCCTGCCGGCTATTG 73
Db 2 GGGTCTCTCGAGGCCCGGCATCTTGACGCTTCAAAAGCGCACGTCCTGCCGGCTGTTTC 61
Qy 74 TGAGCGCTCACAAATTCGGGCTTTTCGACCT 104
Db 62 TCCTCTTCTCATCTCCGGGCTTTTCGACCT 92

RESULT 6
BUS59095
LOCUS
DEFINITION
AGENCOURT 10279964 NIH_MGC_144 Mus musculus cDNA clone
IMAGE:6587745 5', mRNA sequence.
ACCESSION
BUS59095
VERSION
BUS59095.1 GI:22909391
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 798)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.

```

Roberto Alberto, Marianne Beetz-Sargent, Maura Busby, Peter Hunt,
Linda McKinnel, BF Koop.
bioinformatics:
Gordon D Brown.

FEATURES

Location/Qualifiers

1. .697
/organism="Salmo salar"
/mol_type="mRNA"
/strain="McConnell"
/db_xref="taxon:8030"
/clone_lib="whole"
/note="Vector: pBlueScriptIIK; Library Creator: Matthew
L Rise; Atlantic salmon tissue contributors: Carlo Biagi,
Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon
Jones (PBS, Nanaimo, B.C.), Seaspring Hatchery (Crofton,
B.C.), Rachel Roper (University of Victoria)"

ORIGIN

Query Match 52.6%; Score 62.6; DB 13; Length 697;
Best Local Similarity 76.2%; Pred. No. 1.2e-09;
Matches 77; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 19 CTCACAAATCCGGGCTTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGC 78
Db CTTCCGAGCGCGGCTTCGACGCTTCAAAGCGCAGCTCTCGCGGCTTCCTCCTC 553
QY 79 GGTCAAAATCCGGGCTTCGACCTGCGAGCAATATGGGA 119
Db TTTCTCATCTCCGGGCTTCGACCTGCGAGCAATATGGGA 594

RESULT 2

CB237186

LOCUS

DEFINITION CB237186 800 bp mRNA linear EST 10-FEB-2003
IMAGE:30254470 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB237186 800 bp mRNA linear EST 10-FEB-2003
AGENCOURT 11477704 NIH_MGC_166 Mus musculus cDNA clone
IMAGE:30254470 5', mRNA sequence.
EST.
CB237186.1 GI:28288764
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 800)
NTH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDCM24 row: g column: 23
High quality sequence stop: 483.
Location/Qualifiers

FEATURES

source

1. 800
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clones="IMAGE:30254470"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_166"
/note="Organ: brain; Vector: pDNR-LIB; Site 1: SfiI
(ggccattagggc); Site 2: SfiI (ggccgctggggc);
Non-normalized full-length enriched library 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 2.05

ORIGIN

Query Match 41.2%; Score 49; DB 14; Length 800;
Best Local Similarity 71.9%; Pred. No. 3.4e-05;
Matches 64; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 16 GCCTCACATCCGGCATTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTG 75
Db 1 GGCTCGGAGGCCGCGCATTCGACGCTTCAAAGCGCAGCTCTCGCGGCTATTGTG 60
QY 76 AGCGCTCACAAATCCGGGCTTCGACCT 104
Db CTCTTCTCATCTCCGGGCTTCGACCT 89

RESULT 3

CB237098

LOCUS

DEFINITION CB237098 978 bp mRNA linear EST 10-FEB-2003
AGENCOURT 11476567 NIH_MGC_166 Mus musculus cDNA clone
IMAGE:30254413 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CB237098 978 bp mRNA linear EST 10-FEB-2003
AGENCOURT 11476567 NIH_MGC_166 Mus musculus cDNA clone
IMAGE:30254413 5', mRNA sequence.
EST.
CB237098.1 GI:28288676
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 978)
NTH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDCM24 row: e column: 14
High quality sequence stop: 593.
Location/Qualifiers

FEATURES

source

1. 978
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clones="IMAGE:30254413"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_166"
/note="Organ: brain; Vector: pDNR-LIB; Site 1: SfiI
(ggccattagggc); Site 2: SfiI (ggccgctggggc);
Non-normalized full-length enriched library 5' and 3'
adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 2.05
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA).Corp."

ORIGIN

Query Match 41.2%; Score 49; DB 14; Length 978;
Best Local Similarity 71.9%; Pred. No. 3.7e-05;
Matches 64; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 16 GCGCTCACATCCGGCATTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTG 75

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:25:56 ; Search time 726.893 Seconds
(without alignments)
4888.751 Million cell updates/sec

Title: US-09-954-483B-2
Perfect score: 119
Sequence: 1 aagctctattgtgagcgt.....gacctgagccaatatggga 119

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	62.6	52.6	697	13	CA038585 ssalnhw00
2	49	41.2	800	14	CB237186 AGENCOURT
3	49	41.2	978	14	CB237098 AGENCOURT
4	48.2	40.5	1023	13	BUS25591 AGENCOURT

5	47.8	40.2	752	13	BUS71341
6	47.8	40.2	798	13	BUS59095
7	47.6	40.0	168	28	CC199910
8	47.6	40.0	267	13	BY294037
9	47.6	40.0	310	13	BY066197
10	47.6	40.0	326	13	BY124531
11	47.6	40.0	346	13	BY130145
12	47.6	40.0	350	13	BY085570
13	47.6	40.0	353	13	BY163437
14	47.6	40.0	354	13	BY068766
15	47.6	40.0	354	13	BY308363
16	47.6	40.0	358	13	BY074204
17	47.6	40.0	365	13	BY084662
18	47.6	40.0	365	13	BY162545
19	47.6	40.0	376	13	BY292706
20	47.6	40.0	380	13	BY095861
21	47.6	40.0	389	13	BY282249
22	47.6	40.0	399	13	BY161413
23	47.6	40.0	423	13	BY295180
24	47.6	40.0	430	13	BY252840
25	47.6	40.0	433	13	BY274289
26	47.6	40.0	456	13	BY260460
27	47.6	40.0	458	13	BY282583
28	47.6	40.0	464	13	BY254735
29	47.6	40.0	472	13	BY256206
30	47.6	40.0	486	10	BB855320
31	47.6	40.0	499	13	BY244062
32	47.6	40.0	509	14	CF163808
33	47.6	40.0	768	14	CB234230
34	47.6	40.0	768	14	CB601416
35	47.6	40.0	769	9	AU079940
36	47.6	40.0	776	14	CB953941
37	47.6	40.0	777	14	CB234346
38	47.6	40.0	784	13	BUS60967
39	47.6	40.0	790	14	CB236069
40	47.6	40.0	806	14	CB236515
41	47.6	40.0	817	12	BI734054
42	47.6	40.0	848	13	BUS25057
43	47.6	40.0	931	9	AU078897
44	47.6	40.0	941	9	AU067551
45	47.6	40.0	942	14	CB234323

ALIGNMENTS

RESULT 1
CA038585 697 bp mRNA linear EST 04-MAR-2003
LOCUS ssalnhw008036 whole Salmo salar cDNA, mRNA sequence.
DEFINITION CA038585
ACCESSION CA038585.1 GI:24338581
VERSION EST.
KEYWORDS Salmo salar (Atlantic salmon)
SOURCE Salmo salar
ORGANISM Salmo salar
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.

REFERENCE 1 (bases 1 to 697)
AUTHORS GRASP Consortium, Davidson, W.S., Koop, B.F. and
TITLE A survey of Salmo salar transcripts from high complexity cDNA
JOURNAL GRASP web site, <http://web.uvic.ca/cbr/grasp>.
COMMENT A survey of Salmo salar transcripts from high complexity cDNA
Unpublished (2002)
Contact: Koop BF
Centre for Biomedical Research
University of Victoria
PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada
Tel: 250 472 4067
Fax: 250 472 4075
Email: bkoop@uvic.ca
Centre for Biomedical Research, University of Victoria
cDNA preparation and sequencing:

; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-816-790-2

Query Match 52.0%; Score 73.8; DB 9; Length 6355;
Best Local Similarity 87.1%; Pred. No. 1.6e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGCAAGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTGAGCGCTCAAA 86
Db 4625 CCGGGCATTCTCGACGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTCTCTCTCAT 4684

QY 87 TTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 119
Db 4685 CTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 4717

RESULT 14

US-09-861-077-2
; Sequence 2, Application US/09861077
; Patent No. US20020023275A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MATRIX
; FILE REFERENCE: R-15
; CURRENT APPLICATION NUMBER: US/09/861,077
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/204,972
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/215,394
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-861-077-2

Query Match 62.0%; Score 73.8; DB 9; Length 6355;
Best Local Similarity 87.1%; Pred. No. 1.6e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGCAAGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTGAGCGCTCAAA 86
Db 4625 CCGGGCATTCTCGACGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTCTCTCTCAT 4684

QY 87 TTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 119
Db 4685 CTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 4717

RESULT 15

US-09-815-825-2
; Sequence 2, Application US/09815825
; Patent No. US20020026652A1

; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CGMP
; FILE REFERENCE: R-849
; CURRENT APPLICATION NUMBER: US/09/815,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,142
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,227
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/216,765
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/219,182
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-825-2

Query Match 62.0%; Score 73.8; DB 9; Length 6355;
Best Local Similarity 87.1%; Pred. No. 1.6e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGCAAGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTGAGCGCTCAAA 86
Db 4625 CCGGGCATTCTCGACGCTTCAAAAGCGCACGCTCTGCCGGCTATTGTCTCTCTCAT 4684

QY 87 TTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 119
Db 4685 CTCGGGCGCTTTCGACCTGCGAGCAATATGGGA 4717

Search completed: April 22, 2004, 13:16:13
Job time : 107.145 secs

```
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,348
; PRIOR FILING DATE: 2000-03-16
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage Vector
US-09-811-361-1

Query Match      62.0%; Score 73.8; DB 13; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86
    |||||
Db 3038 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86

QY 87 TTCGGGCGCTTTGACCTGCGAGCAATATGGGA 119
    |||||
Db 3098 CTCGGGCGCTTTGACCTGCGAGCAATATGGGA 3130

RESULT 10
US-09-815-937-1
; Sequence 1, Application US/09815937
; Publication No. US2002008012A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: LYMPHOID-SPECIFIC GPCR GENE DISRUPTIONS
; FILE REFERENCE: R-611
; CURRENT APPLICATION NUMBER: US/09/815,937
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,128
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/221,485
; PRIOR FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-937-1

Query Match      62.0%; Score 73.8; DB 13; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86
    |||||
Db 3038 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86

QY 87 TTCGGGCGCTTTGACCTGCGAGCAATATGGGA 119
    |||||
Db 3098 CTCGGGCGCTTTGACCTGCGAGCAATATGGGA 3130

RESULT 11
US-09-885-816-1
; Sequence 1, Application US/09885816
; Publication No. US20020086369A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
US-09-816-790-2
; Sequence 2, Application US/09816790
; Patent No. US2002002255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
US-10-087-523-1
; Sequence 1, Application US/10087523
; Publication No. US20020197624A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING
; TITLE OF INVENTION: SEQUENCES INTO EMBRYONIC STEM CELLS
; FILE REFERENCE: 376472000200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-11-17
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Plasmid vector
US-10-087-523-1

Query Match      62.0%; Score 73.8; DB 14; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86
    |||||
Db 3038 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86

QY 87 TTCGGGCGCTTTGACCTGCGAGCAATATGGGA 119
    |||||
Db 3098 CTCGGGCGCTTTGACCTGCGAGCAATATGGGA 3130

RESULT 12
US-10-087-523-1
; Sequence 1, Application US/10087523
; Publication No. US20020197624A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING
; TITLE OF INVENTION: SEQUENCES INTO EMBRYONIC STEM CELLS
; FILE REFERENCE: 376472000200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-11-17
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Plasmid vector
US-10-087-523-1

Query Match      62.0%; Score 73.8; DB 14; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86
    |||||
Db 3038 CCGGGCATTCTCGAAGCTTCAAAAGCGCAGCTCTGCGGCTATTGTGAGCGCTCACAA 86

QY 87 TTCGGGCGCTTTGACCTGCGAGCAATATGGGA 119
    |||||
Db 3098 CTCGGGCGCTTTGACCTGCGAGCAATATGGGA 3130

RESULT 13
US-09-816-790-2
; Sequence 2, Application US/09816790
; Patent No. US2002002255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
```


Qy	27	CCCGGCA	TTCTCGCAGC	TTCAAAAGCG	CACGTC	TGCGCG	CGCTATT	TGTGAG	CGCGTCACAA	86
Db	3038	CCCGGCA	TTCTCGCAGC	TTCAAAAGCG	CACGTC	TGCGCG	CGCTATT	TCTCCT	TTCCGTCAT	3097
Qy	87	TTCCGGCG	CGCTTCGAC	CTGCAG	CGCAATAT	CGGA				119
Db	3098	CTCCGGCG	CGCTTCGAC	CTGCAG	CGCAATAT	CGGA				3130

```

RESULT 6
US-09-815-825-1
/ Sequence 1, Application US/09815825
/ Patent No. US20020026652A1
/ GENERAL INFORMATION:
/ APPLICANT: Allen, Keith D.
/ APPLICANT: Phillips, Russell
/ TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CGMP
/ TITLE OF INVENTION: PHOSPHODIESTERASE GENE DISRUPTIONS
/ FILE REFERENCE: R-849
/ CURRENT APPLICATION NUMBER: US/09/815,825
/ CURRENT FILING DATE: 2001-03-22
/ PRIOR APPLICATION NUMBER: US 60/191,142
/ PRIOR FILING DATE: 2000-03-22
/ PRIOR APPLICATION NUMBER: US 60/204,227
/ PRIOR FILING DATE: 2000-05-15
/ PRIOR APPLICATION NUMBER: US 60/216,765
/ PRIOR FILING DATE: 2000-07-06
/ PRIOR APPLICATION NUMBER: US 60/219,182
/ PRIOR FILING DATE: 2000-07-19
/ NUMBER OF SEQ ID NOS: 21
/ SOFTWARE: FastSEQ for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 4768
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Phase vector
US-09-815-825-1

```

	Query Match	62.08;	Score 73.8;	DB 9;	Length 4758;
	Best Local Similarity	87.11;	Pred. No. 1.5e-17;		
	Matches 81;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
Qy	27	CCCGGCATTTCGCAACCTTCAAAGCGCACGCTGCGCGCTATTGTGAGCGCTCAAA	86		
Db	3038	CCCGGCATTTCGCAACCTTCAAAGCGCACGCTGCGCGCTATTGTCTCTCTCTCAT	3097		
Qy	87	TTCCGGGSCCTTTTCGACCTGCAGCCAAATATGGGA	119		
Db	3098	CTCCGGGSCCTTTTCGACCTGCAGCCAAATATGGGA	3130		

```

RESULT 7
US-09-815-935-1
; Sequence 1, Application US/09815935
; Patent No. US20020038466A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MAGNESIUM
; TITLE OF INVENTION: DEPENDENT PROTEIN PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-723
; CURRENT APPLICATION NUMBER: US/09/815,935
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,235
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/216,249
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA

```

```

; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Phage vector
US-09-815-935-1

```

	Query Match	62.0%;	Score 73.8;	DB 9;	Length 4768;
	Best Local Similarity	87.8%;	Pred. No. 1.5e-17;		
	Matches 81;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
Qy	27	CCCGGATTCTCGCAAGCTTCAAAGGCGACGCTGTGCGGCGCTATTGTGTGAGCGCTCACAA	86		
Db	3038	CCCGGATTCTCGACGCTTCAAAGGCGACGCTGTGCGGCGTTCCTCTCTCTCAT	3097		
Qy	87	TTCCGGGCGCTTTTCGACGCTGCAGCCAATATGGGA	119		
Db	3098	CTCCGGGCGCTTTTCGACGCTGCAGCCAATATGGGA	3130		

```

RESULT 8
US-09-815-944-1
; Sequence 1, Application US/09815944
; Patent No. US20020038467A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MELANOCYTE
; TITLE OF INVENTION: STIMULATING HORMONE RECEPTOR GENE DISRUPTIONS
; FILE REFERENCE: R-654
; CURRENT APPLICATION NUMBER: US/09/815,944
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,236
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/215,214
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/218,075
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/219,167
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phase vector
; US-09-815-944-1

```

	Query Match	62.0%	Score 73.8	DB 9	Length 4768
	Best Local Similarity	87.1%	Pred. NO. 1.5e-17		
	Matches 81	Conservative 0	Mismatches 12	Indels 0	Gaps 0
Qy	27	CCCGGCATTCTCCAAAGCTTCAAAGCGCACGCTCTCCGCGCTATTGTGAGCGCTCAAA	86		
Db	3038	CCCGGCATTCTCGACGCTTCAAAGCGCACGCTCTCCGCGCTGTTCTCTCTCTCAT	3097		
Qy	87	TTCCGGGCGCTTTTCGACCTGCGACCAATATGGGA	119		
Db	3098	CTCCGGGCGCTTTTCGACCTGCGACCAATATGGGA	3130		

RESULT 9
US-09-811-361-1
; Sequence 1, Application US/09811361
; Publication No. US20040040046A1
; GENERAL INFORMATION:
; APPLICANT: Guenther, Catherine
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: RETINA-SPECIFIC NUCLEAR RECEPTOR GENE DISRUPTIONS
; FILE REFERENCE: R-125
; CURRENT APPLICATION NUMBER: US/09/811.361

```

RESULT 2
US-09-954-483A-6
; Sequence 6, Application US/09954483A
; Publication No. US20030032175A1
; GENERAL INFORMATION:
; APPLICANT: Siebel, Christian
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF PRODUCING CELLS AND ANIMALS
; TITLE OF INVENTION: COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: RMES-02
; CURRENT APPLICATION NUMBER: US/09/954,483A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/232,957
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-09-954-483A-6

Query Match      63.0%; Score 75; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 2.2e-18;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 GCATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTGAGCGCTCACAATCC 90
DB 1 GCATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTGAGCGCTCACAATCC 60

QY 91 GGGCCCTTCGACCTG 105
DB 61 GGGCCCTTCGACCTG 75

RESULT 3
US-09-954-483A-1
; Sequence 1, Application US/09954483A
; Publication No. US20030032175A1
; GENERAL INFORMATION:
; APPLICANT: Siebel, Christian
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF PRODUCING CELLS AND ANIMALS
; TITLE OF INVENTION: COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: RMES-02
; CURRENT APPLICATION NUMBER: US/09/954,483A
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/232,957
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 108
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Targeting Vector
US-09-954-483A-1

Query Match      62.0%; Score 73.8; DB 10; Length 108;
Best Local Similarity 87.1%; Pred. No. 6.9e-18;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCCGGCATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTGAGCGCTCACA 86
DB 16 CCCGGCATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTCTCTCTCAT 75

QY 87 TTCGGGCGCTTTCGACCTCGAGCCAATATGGGA 119

```

```

DB 76 CTCGGGCGCTTTCGACCTCGAGCCAATATGGGA 108

RESULT 4
US-09-816-790-1
; Sequence 1, Application US/09816790
; Patent No. US20020022255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-816-790-1

Query Match      62.0%; Score 73.8; DB 9; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTGAGCGCTCACA 86
DB 3038 CCGGCAATTCTCGCAAGCTTCAAAGCGCACGCTCGCGCGCTATTGTCTCTCTCAT 3097

QY 87 TTCGGGCGCTTTCGACCTCGAGCCAATATGGGA 119
DB 3098 CTCGGGCGCTTTCGACCTCGAGCCAATATGGGA 3130

RESULT 5
US-09-861-077-1
; Sequence 1, Application US/09861077
; Patent No. US20020023275A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MATRIX
; TITLE OF INVENTION: METALLOPROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-15
; CURRENT APPLICATION NUMBER: US/09/861,077
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/204,972
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/215,394
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-861-077-1

Query Match      62.0%; Score 73.8; DB 9; Length 4768;
Best Local Similarity 87.1%; Pred. No. 1.5e-17;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:42:41 ; Search time 103.145 Seconds
(without alignments)
5201.676 Million cell updates/sec

Title: US-09-954-483B-2
Perfect score: 119
Sequence: 1 aaggtctattgtgagcgtc.....gacctgcagcaatatggga 119

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues
Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubna/US06_PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubna/US09_PCT_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubna/US10C_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubna/US10C_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	119	10	US-09-954-483A-2
2	75	63.0	75	10	US-09-954-483A-6
3	73.8	62.0	108	10	US-09-954-483A-1
4	73.8	62.0	4768	9	US-09-816-790-1
5	73.8	62.0	4768	9	US-09-861-077-1
6	73.8	62.0	4768	9	US-09-815-825-1
7	73.8	62.0	4768	9	US-09-815-935-1
8	73.8	62.0	4768	9	US-09-815-944-1
9	73.8	62.0	4768	13	US-09-811-361-1
10	73.8	62.0	4768	13	US-09-815-937-1
11	73.8	62.0	4768	13	US-09-885-816-1
12	73.8	62.0	4768	14	US-10-087-523-1
13	73.8	62.0	6355	9	US-09-816-790-2
14	73.8	62.0	6355	9	US-09-861-077-2

15	73.8	62.0	6355	9	US-09-815-825-2	Sequence 2, Appli
16	73.8	62.0	6355	9	US-09-815-935-2	Sequence 2, Appli
17	73.8	62.0	6355	13	US-09-815-944-2	Sequence 2, Appli
18	73.8	62.0	6355	13	US-09-811-361-2	Sequence 2, Appli
19	73.8	62.0	6355	13	US-09-815-937-2	Sequence 2, Appli
20	73.8	62.0	6355	13	US-09-885-816-2	Sequence 2, Appli
21	73.8	62.0	6355	14	US-10-087-523-2	Sequence 2, Appli
22	67.8	57.0	538	13	US-10-614-116-3	Sequence 3, Appli
23	63.4	53.3	8934	10	US-09-843-150-56	Sequence 56, Appli
24	62.6	52.6	10491	15	US-10-359-050-18	Sequence 18, Appli
25	62.6	52.6	12538	15	US-10-359-050-12	Sequence 12, Appli
26	62.6	52.6	12545	15	US-10-359-050-13	Sequence 13, Appli
27	56.8	47.7	11784	15	US-10-359-050-20	Sequence 20, Appli
28	56.8	47.7	11784	16	US-10-014-099F-106	Sequence 106, Appli
29	56	47.1	66	10	US-09-954-483A-8	Sequence 8, Appli
30	53.6	45.0	7523	16	US-10-014-099F-79	Sequence 79, Appli
31	53.6	45.0	7608	16	US-10-014-099F-78	Sequence 78, Appli
32	53.6	45.0	7803	16	US-10-014-099F-82	Sequence 82, Appli
33	53.6	45.0	8167	16	US-10-014-099F-83	Sequence 83, Appli
34	53	44.5	5382	15	US-10-331-329-21	Sequence 21, Appli
35	53	44.5	9737	15	US-10-331-329-22	Sequence 22, Appli
36	53	44.5	9737	15	US-10-331-329-23	Sequence 23, Appli
37	53	44.5	9737	15	US-10-331-329-28	Sequence 28, Appli
38	53	44.5	9871	15	US-10-331-329-24	Sequence 24, Appli
39	53	44.5	10060	15	US-10-331-329-25	Sequence 25, Appli
40	50.6	42.5	17135	10	US-09-843-150-55	Sequence 55, Appli
41	48.6	40.8	9027	15	US-10-117-960-11	Sequence 11, Appli
42	48.6	40.8	10551	15	US-10-117-960-1	Sequence 1, Appli
43	48.6	40.8	11646	15	US-10-117-960-10	Sequence 10, Appli
44	48.6	40.8	12041	15	US-10-117-960-9	Sequence 9, Appli
45	48.6	40.8	13547	15	US-10-117-960-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-954-483A-2
; Sequence 2, Application US/09954483A
; Publication No. US20030032175A1
; GENERAL INFORMATION: Siebel, Christian J.
; APPLICANT: Siebel, Christian J.
; TITLE OF INVENTION: METHODS OF PRODUCING CELLS AND ANIMALS
; TITLE OF INVENTION: COMPRISING TARGETED GENE MODIFICATIONS AND COMPOSITIONS
; TITLE OF INVENTION: RELATING THERETO
; FILE REFERENCE: RMES-02
; CURRENT APPLICATION NUMBER: US/09/954,483A
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US 60/232,957
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 119
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Targeting Vector
US-09-954-483A-2

Query Match 100.0%; Score 119; DB 10; Length 119;
Best Local Similarity 100.0%; Pred. No. 5.7e-35;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGTCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 60
1 AAGTCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 60
Db 1 AAGTCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 60
QY 61 TCGCCGCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 119
61 TCGCCGCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 119
Db 61 TCGCCGCTATTGTGAGCGCTCACAATCCCGGCTTCGCAAGCTTCAAAGCGCAGTC 119

; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-28

Query Match 44.5%; Score 53; DB 4; Length 9737;
Best Local Similarity 73.1%; Pred. No. 6.6e-10; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 19 CTCACAATCCGGCATTCTCGAAGCTTCAAAGCGCACGTCGCGGCTATTGTGAGC 78
Db 8563 CTCGGAGGCCCGGCATTCTGCACGCTTCAAAGCGCACGTCGCGGCTATTGTGAGC 78

QY 79 GCTCACAAATTCGGGCGCTTTGACCTGCAGCCA 111
Db 8623 TTCTCATCTCCGGGCGCTTTGACCTGCATCCA 8655

RESULT 15
US-09-484-997-22
; Sequence 22, Application US/09484997
; Patent No. 6524818
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERP, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003J
; CURRENT APPLICATION NUMBER: US/09/484,997
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-22

Query Match 44.5%; Score 53; DB 4; Length 9737;
Best Local Similarity 73.1%; Pred. No. 6.6e-10; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 19 CTCACAATCCGGCATTCTCGAAGCTTCAAAGCGCACGTCGCGGCTATTGTGAGC 78
Db 8563 CTCGGAGGCCCGGCATTCTGCACGCTTCAAAGCGCACGTCGCGGCTATTGTGAGC 78

QY 79 GCTCACAAATTCGGGCGCTTTGACCTGCAGCCA 111
Db 8623 TTCTCATCTCCGGGCGCTTTGACCTGCATCCA 8655

Search completed: April 22, 2004, 12:58:13
Job time : 25.082 secs

QY 19 CTCACATCCCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGC 78
Db 1106 CTCGGAGGCGCGCATTCGACGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTCTCTC 1165
QY 79 GCTCACATTCGGGCGCTTTCGACCTGCAGCCA 111
Db 1166 TTCCTCATCTCCGGGCGCTTTCGACCTGCATCCA 1198

RESULT 12

US-09-479-122-22
; Sequence 22, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; PRIOR FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-22

Query Match 44.5%; Score 53; DB 4; Length 9737;

Best Local Similarity 73.1%; Pred. No. 6.6e-10; Mismatches 25; Indels 0; Gaps 0;

QY 19 CTCACATCCCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGC 78
Db 8563 CTCGGAGGCGCGCATTCGACGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTCTCTC 8622
QY 79 GCTCACATTCGGGCGCTTTCGACCTGCAGCCA 111
Db 8623 TTCCTCATCTCCGGGCGCTTTCGACCTGCATCCA 8655

RESULT 13

US-09-479-122-23
; Sequence 23, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122

; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 23
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-23

Query Match 44.5%; Score 53; DB 4; Length 9737;

Best Local Similarity 73.1%; Pred. No. 6.6e-10; Mismatches 25; Indels 0; Gaps 0;

QY 19 CTCACATCCCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGC 78
Db 8563 CTCGGAGGCGCGCATTCGACGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTCTCTC 8622
QY 79 GCTCACATTCGGGCGCTTTCGACCTGCAGCCA 111
Db 8623 TTCCTCATCTCCGGGCGCTTTCGACCTGCATCCA 8655

RESULT 14

US-09-479-122-28
; Sequence 28, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base

```
Db 1166 TTCTCATCTCCGGGCGCTTTTCGACCTGCATCCA 1198

RESULT 9
US-09-455-659A-21
; Sequence 21, Application US/09455659A
; Patent No. 6602686
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003A
; CURRENT APPLICATION NUMBER: US/09/455,659A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-484-996-21
; Query Match 44.5%; Score 53; DB 4; Length 5382;
; Best Local Similarity 73.1%; Pred. No. 5.5e-10;
; Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 19 CTCACAATCCGGGCGCTTTTCGACCTGCATCCA 111
|||
Db 1106 CTCGGAGGCCGCGCATTTCTGCACGCTTCAAAGCGCACGCTCTCCCGCTATTGTGAGC 78
|||

Qy 79 GCTCACAATCCGGGCGCTTTTCGACCTGCATCCA 111
|||
Db 1166 TTCTCATCTCCGGGCGCTTTTCGACCTGCATCCA 1198
|||

RESULT 11
US-09-479-123-21
; Sequence 21, Application US/09479123
; Patent No. 6670185
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003B
; CURRENT APPLICATION NUMBER: US/09/479,123
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
; US-09-479-123-21
; Query Match 44.5%; Score 53; DB 4; Length 5382;
; Best Local Similarity 73.1%; Pred. No. 5.5e-10;
; Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 19 CTCACAATCCGGGCGCTTTTCGACCTGCATCCA 111
|||
Db 1106 CTCGGAGGCCGCGCATTTCTGCACGCTTCAAAGCGCACGCTCTCCCGCTATTGTGAGC 78
|||

Qy 79 GCTCACAATCCGGGCGCTTTTCGACCTGCATCCA 111
|||
Db 1166 TTCTCATCTCCGGGCGCTTTTCGACCTGCATCCA 1198
|||

RESULT 10
US-09-484-996-21
; Sequence 21, Application US/09484996
; Patent No. 6623958
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003H
; CURRENT APPLICATION NUMBER: US/09/484,996
; CURRENT FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1998-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
```

Patent No. 6524818
GENERAL INFORMATION:
APPLICANT: HARRINGTON, JOHN J.
APPLICANT: SHERP, BRUCE
APPLICANT: RUNDLETT, STEPHEN
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
TITLE OF INVENTION: ENDOGENOUS GENES
FILE REFERENCE: 0221-0003J
CURRENT FILING DATE: 2000-01-18
PRIORITY APPLICATION NUMBER: 09/276,820
PRIORITY FILING DATE: 1999-03-26
PRIORITY APPLICATION NUMBER: 09/159,643
PRIORITY FILING DATE: 1998-09-24
PRIORITY APPLICATION NUMBER: 08/941,223
PRIORITY FILING DATE: 1997-09-26
PRIORITY APPLICATION NUMBER: 09/263,814
PRIORITY FILING DATE: 1999-03-08
PRIORITY APPLICATION NUMBER: 09/253,022
PRIORITY FILING DATE: 1999-02-19
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 21
LENGTH: 5382
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (890)
OTHER INFORMATION: a, c, t, g, other or unknown
NAME/KEY: modified_base
LOCATION: (1042)
OTHER INFORMATION: a, c, t, g, other or unknown
US-09-484-997-21

Query Match 44.5%; Score 53; DB 4; Length 5382;
Best Local Similarity 73.1%; Pred. No. 5.5e-10;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 19 CTCACAATCCCGGCATCTCGCAAGCTTCAAAAGCGCAGCTCGCGCGCTATTGTGAGC 78
Db 1106 CTCGGAGGCCGCGCATCTGACGCTTCAAAAGCGCAGCTCGCGCGCTATTGTCTCTC 1165

QY 79 GCTCACAATCCCGGCCTTTGACCTGCAGCCA 111
Db 1166 TTCCTCATCTCCGGCGCTTTCGACCTGCATCCA 1198

RESULT 7
US-09-481-355-21
Sequence 21, Application US/09481355
Patent No. 6524824
GENERAL INFORMATION:
APPLICANT: HARRINGTON, JOHN J.
APPLICANT: SHERP, BRUCE
APPLICANT: RUNDLETT, STEPHEN
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
TITLE OF INVENTION: ENDOGENOUS GENES
FILE REFERENCE: 0221-0003F
CURRENT FILING DATE: 2000-01-12
PRIORITY APPLICATION NUMBER: 09/276,820
PRIORITY FILING DATE: 1999-03-26
PRIORITY APPLICATION NUMBER: 09/159,643
PRIORITY FILING DATE: 1998-09-24
PRIORITY APPLICATION NUMBER: 08/941,223
PRIORITY FILING DATE: 1997-09-26
PRIORITY APPLICATION NUMBER: 09/263,814
PRIORITY FILING DATE: 1999-03-08
PRIORITY APPLICATION NUMBER: 09/253,022
PRIORITY FILING DATE: 1999-02-19
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 21
LENGTH: 5382
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (890)
OTHER INFORMATION: a, c, t, g, other or unknown
NAME/KEY: modified_base
LOCATION: (1042)
OTHER INFORMATION: a, c, t, g, other or unknown
US-09-481-355-21

Query Match 44.5%; Score 53; DB 4; Length 5382;
Best Local Similarity 73.1%; Pred. No. 5.5e-10;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 19 CTCACAATCCCGGCATCTCGCAAGCTTCAAAAGCGCAGCTCGCGCGCTATTGTGAGC 78
Db 1106 CTCGGAGGCCGCGCATCTGACGCTTCAAAAGCGCAGCTCGCGCGCTATTGTCTCTC 1165
QY 79 GCTCACAATCCCGGCCTTTGACCTGCAGCCA 111
Db 1166 TTCCTCATCTCCGGCGCTTTCGACCTGCATCCA 1198

RESULT 8
US-09-481-282-21
Sequence 21, Application US/09481282
Patent No. 6541221
GENERAL INFORMATION:
APPLICANT: HARRINGTON, JOHN J.
APPLICANT: SHERP, BRUCE
APPLICANT: RUNDLETT, STEPHEN
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
TITLE OF INVENTION: ENDOGENOUS GENES
FILE REFERENCE: 0221-0003US
CURRENT FILING DATE: 2000-01-11
PRIORITY APPLICATION NUMBER: 09/159,643
PRIORITY FILING DATE: 1998-09-24
PRIORITY APPLICATION NUMBER: 08/941,223
PRIORITY FILING DATE: 1997-09-26
PRIORITY APPLICATION NUMBER: 09/263,814
PRIORITY FILING DATE: 1999-03-08
PRIORITY APPLICATION NUMBER: 09/253,022
PRIORITY FILING DATE: 1999-02-19
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 21
LENGTH: 5382
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (890)
OTHER INFORMATION: a, c, t, g, other or unknown
NAME/KEY: modified_base
LOCATION: (1042)
OTHER INFORMATION: a, c, t, g, other or unknown
US-09-481-282-21

Query Match 44.5%; Score 53; DB 4; Length 5382;
Best Local Similarity 73.1%; Pred. No. 5.5e-10;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 19 CTCACAATCCCGGCATCTCGCAAGCTTCAAAAGCGCAGCTCGCGCGCTATTGTGAGC 78
Db 1106 CTCGGAGGCCGCGCATCTGACGCTTCAAAAGCGCAGCTCGCGCGCTATTGTCTCTC 1165
QY 79 GCTCACAATCCCGGCCTTTGACCTGCAGCCA 111

```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-08-693-174-4

Query Match      49.4%; Score 58.8; DB 2; Length 1110;
Best Local Similarity 84.6%; Pred. No. 2.3e-12;
Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86
Db 853 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86

QY 87 TTCGGGCGCTTTTCGACCT 104
Db 913 CTCGGGCGCTTTTCGACCT 930

RESULT 3
US-09-738-4
; Sequence 4, Application US/09253738
; Patent No. 6265390
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewelyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/09/253,738
; CURRENT FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-09-253-738-4

Query Match      49.4%; Score 58.8; DB 3; Length 1110;
Best Local Similarity 84.6%; Pred. No. 2.3e-12;
Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86
Db 853 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86

QY 87 TTCGGGCGCTTTTCGACCT 104
Db 913 CTCGGGCGCTTTTCGACCT 930

RESULT 4
US-09-714-550-1
; Sequence 1, Application US/09714550
; Patent No. 6558948
; GENERAL INFORMATION:
; APPLICANT: Kochanek, Stefan
; APPLICANT: Schiedner, Gudrun
; TITLE OF INVENTION: Permanent amniocytic cell line, its
; TITLE OF INVENTION: production and use for the production of gene transfer
; TITLE OF INVENTION: vectors
; FILE REFERENCE: 50125/007002
; CURRENT APPLICATION NUMBER: US/09/714,550
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/167,439
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0

US-09-954-483b-2.rni

; SEQ ID NO 1
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-714-550-1

Query Match      48.6%; Score 57.8; DB 4; Length 513;
Best Local Similarity 84.4%; Pred. No. 4.3e-12;
Matches 65; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86
Db 437 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGCGCTCAAA 86

QY 87 TTCGGGCGCTTTTCGACCT 103
Db 497 CTCGGGCGCTTTTCGACCT 513

RESULT 5
US-09-479-122-21
; Sequence 21, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-21

Query Match      44.5%; Score 53; DB 4; Length 5382;
Best Local Similarity 73.1%; Pred. No. 5.5e-10;
Matches 68; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 19 CTCACAATCCGGCATTCTCGAAGCTTTCGACCTTCGACCTGCAGCCA 111
Db 1106 CTCGGAGGCCCGCATTCTTCGACCTTCAAAAGCGCAGCTCTGCGGCGCTATTGTGAGC 78

QY 79 GCTCACAATTCGGGCGCTTTTCGACCTTCGACCTGCAGCCA 111
Db 1166 TTCTCATCTCCGGGCGCTTTTCGACCTTCGACCTGCATCCA 1198

RESULT 6
US-09-484-997-21
; Sequence 21, Application US/09484997
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:26:21 ; Search time 21.082 Seconds
(without alignments)
3132.491 Million cell updates/sec

Title: US-09-954-483B-2
Perfect score: 119
Sequence: 1 aagctctattgagcgct.....gacctgagccaatatggga 119

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52.2	49.7	7090	4	US-09-714-550-18
2	58.8	49.4	1110	2	US-08-693-174-4
3	58.8	49.4	1110	3	US-09-253-738-4
4	57.8	48.6	513	4	US-09-714-550-1
5	53	44.5	5382	4	US-09-479-122-21
6	53	44.5	5382	4	US-09-484-997-21
7	53	44.5	5382	4	US-09-481-355-21
8	53	44.5	5382	4	US-09-481-282-21
9	53	44.5	5382	4	US-09-455-659A-21
10	53	44.5	5382	4	US-09-484-996-21
11	53	44.5	5382	4	US-09-479-123-21
12	53	44.5	9737	4	US-09-479-122-22
13	53	44.5	9737	4	US-09-479-122-23
14	53	44.5	9737	4	US-09-479-122-28
15	53	44.5	9737	4	US-09-484-997-22
16	53	44.5	9737	4	US-09-484-997-23
17	53	44.5	9737	4	US-09-484-997-28
18	53	44.5	9737	4	US-09-481-355-22
19	53	44.5	9737	4	US-09-481-355-23
20	53	44.5	9737	4	US-09-481-355-28
21	53	44.5	9737	4	US-09-481-282-22
22	53	44.5	9737	4	US-09-481-282-23
23	53	44.5	9737	4	US-09-481-282-28
24	53	44.5	9737	4	US-09-455-659A-22
25	53	44.5	9737	4	US-09-455-659A-23
26	53	44.5	9737	4	US-09-455-659A-28
27	53	44.5	9737	4	US-09-484-996-22

28	53	44.5	9737	4	US-09-484-996-23	Sequence 23, Appl
29	53	44.5	9737	4	US-09-484-996-28	Sequence 28, Appl
30	53	44.5	9737	4	US-09-479-123-22	Sequence 22, Appl
31	53	44.5	9737	4	US-09-479-123-23	Sequence 23, Appl
32	53	44.5	9737	4	US-09-479-123-28	Sequence 28, Appl
33	53	44.5	9871	4	US-09-479-122-24	Sequence 24, Appl
34	53	44.5	9871	4	US-09-484-997-24	Sequence 24, Appl
35	53	44.5	9871	4	US-09-481-355-24	Sequence 24, Appl
36	53	44.5	9871	4	US-09-481-282-24	Sequence 24, Appl
37	53	44.5	9871	4	US-09-455-659A-24	Sequence 24, Appl
38	53	44.5	9871	4	US-09-484-996-24	Sequence 24, Appl
39	53	44.5	9871	4	US-09-479-123-24	Sequence 24, Appl
40	53	44.5	10060	4	US-09-479-122-25	Sequence 25, Appl
41	53	44.5	10060	4	US-09-484-997-25	Sequence 25, Appl
42	53	44.5	10060	4	US-09-481-355-25	Sequence 25, Appl
43	53	44.5	10060	4	US-09-481-282-25	Sequence 25, Appl
44	53	44.5	10060	4	US-09-455-659A-25	Sequence 25, Appl
45	53	44.5	10060	4	US-09-484-996-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-714-550-18
; Sequence 18, Application US/09714550
; Patent No. 6558948
; GENERAL INFORMATION:
; APPLICANT: Kochanek, Stefan
; APPLICANT: Schiedner, Gudrun
; TITLE OF INVENTION: Permanent amniocytic cell line, its
; TITLE OF INVENTION: production and use for the production of gene transfer
; TITLE OF INVENTION: vectors
; FILE REFERENCE: 50125/007002
; CURRENT APPLICATION NUMBER: US/09/714,550
; CURRENT FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/167,439
; PRIOR FILING DATE: 1999-11-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 7090
; TYPE: DNA
; ORGANISM: Plasmid STK146
US-09-714-550-18

Query Match 49.7%; Score 59.2; DB 4; Length 7090;
Best Local Similarity 83.8%; Pred. No. 3e-12;
Matches 67; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY	27	CCGCGCATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA	86
Db	2665	CCGCGCATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA	86
QY	87	TTCCGCGCGCTTTTCGACCTGC	106
Db	2725	CTCCGCGCGCTTTTCGACCTGC	2744

RESULT 2
US-08-693-174-4
; Sequence 4, Application US/08693174A
; Patent No. 5942434
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/08/693,174A
; CURRENT FILING DATE: 1996-12-12

ID AAD28660 standard; DNA; 6355 BP.
 AC AAD28660;
 DT 07-MAY-2002 (first entry)
 DE Plasmid pDG4 vector.
 KW Plasmid pDG4; DNA construct; embryonic stem cell, cell disruption; Neo';
 KW neomycin; ampicillin resistance gene; green fluorescent protein; GFP; ds.
 XX Unidentified.
 OS WO200204621-A2.
 PN WO200204621-A2.
 XX 17-JAN-2002.
 PD 11-JUL-2000; 2000WO-US018812.
 PF 11-JUL-2000; 2000WO-US018812.
 PR 11-JUL-2000; 2000WO-US018812.
 XX (DELT-) DELTAGEN INC.
 PA Klein RD, Brennan TJ;
 PI WPI; 2002-164642/21.
 DR Novel nucleotide construct for generating DNA constructs for introducing
 XX into embryonic stem cell, comprising a sequence encoding a positive
 PT selection marker flanked by restriction enzyme sites.
 FT Claim 11; Fig 3B; 64pp; English.
 XX The invention relates to nucleotide construct for generating DNA
 CC constructs. The nucleotide construct comprises a sequence encoding a
 CC positive selection marker flanked by restriction enzyme sites, where
 CC restriction site is flanked by sequences which are not complementary to
 CC each other and which do not include at least one type of base at any
 CC position, where the construct can be treated so that single-stranded
 CC regions are created at each sequence lacking at least one nucleotide. The
 CC nucleotide construct is useful in a rapid and efficient method for
 CC generating DNA constructs suitable for introduction into embryonic stem
 CC cells and for disrupting the function of a gene in a cell. The present
 CC sequence is plasmid pDG4 vector construct containing an ampicillin
 CC resistance gene, a neomycin gene (Neo') and a green fluorescent protein
 CC (GFP) gene. On each site of the Neo' gene are two sites for ligation
 CC independent cloning along with restriction sites
 XX Sequence 6355 BP; 1531 A; 1670 C; 1655 G; 1499 T; 0 U; 0 Other;
 SQ
 Query Match 62.0%; Score 73.8; DB 6; Length 6355;
 Best Local Similarity 87.1%; Pred. No. 7.8e-16;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 27 CCGGCAATCTCGCAAGCTTCAAAAGCGCAGCTGTCGGCGCTATTGTGAGCGCTCAAA 86
 DB 4625 CCGGCAATCTCGCAAGCTTCAAAAGCGCAGCTGTCGGCGCTATTGTGAGCGCTCAAT 4684
 QY 87 TTCGGGCGCTTTCGACCTGCGAGCAATATGGA 119
 DB 4685 CTCGGGCGCTTTCGACCTGCGAGCAATATGGA 4717
 RESULT 14
 ID ADB68453
 AC ADB68453;
 DT 04-DEC-2003 (first entry)
 DE Mouse DCAL knockout targeting vector DNA.
 XX

KW dendritic cell expressed S-adenosyl homocysteine hydrolase-like molecule;
 KW DC; AHCY; DCAL; antiallergic; immunosuppressive; allergy; asthma;
 KW allergic rhinitis; systemic anaphylaxis; autoimmune; diabetes mellitus;
 KW rheumatoid arthritis; transplant rejection; vaccine; gene therapy;
 XX knockout targeting vector; ds.
 OS Synthetic.
 OS Mus sp.
 XX Key Location/Qualifiers
 FT misc_feature 63..5140
 FT /tag= a
 FT /note= "5' transgene (TG) arm"
 FT 5141..7171
 FT /tag= b
 FT /note= "Neo cassette region"
 FT 7171..7275
 FT /tag= c
 FT /note= "Mouse DCAL exon 1 inserted into synthetic
 FT construct"
 FT 7172..8817
 FT /tag= d
 FT /note= "3' transgene (TG) arm"
 FT 8178..8278
 FT /tag= e
 FT /note= "Mouse DCAL exon 2 inserted into synthetic
 FT construct"
 FT 8748..8816
 FT /tag= f
 FT /note= "Mouse DCAL exon 3 inserted into synthetic
 FT construct"
 XX WO2003055997-A1.
 XX 10-JUL-2003.
 XX 24-DEC-2002; 2002WO-AU001761.
 XX 24-DEC-2001; 2001AU-00009741.
 XX (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.
 XX Kato M, Angel NZ, Cooper BJ, Hart DNJ;
 XX WPI; 2003-559275/52.
 PT New dendritic cell expressed S-adenosyl homocysteine hydrolase-like
 PT molecule (DCAL) gene, useful for preparing a composition for treating or
 PT preventing a condition e.g., allergy, autoimmune disease or transplant
 PT rejection.
 XX Example 4; Fig 9; 189pp; English.
 PS The invention relates to a novel isolated polynucleotide comprising a
 CC dendritic cell (DC)-expressed S-adenosyl homocysteine hydrolase (AHCY)-
 CC like molecule (DCAL) gene. The polynucleotide of the invention
 CC demonstrates antiallergic and immunosuppressive activities and may be
 CC useful for treating a condition including an allergy such as asthma,
 CC allergic rhinitis or systemic anaphylaxis, an autoimmune disease such as
 CC diabetes mellitus or rheumatoid arthritis or transplant rejection.
 CC Furthermore, the polynucleotide may be useful as a vaccine or during gene
 CC therapy procedures. The genetically modified animal or the identified
 CC modulatory agent of the invention may be used in the study of immunity,
 CC DC function, brain physiology or neuronal cell function. The current
 CC sequence is that of the mouse DCAL knockout targeting vector DNA of the
 CC invention.
 XX Sequence 13462 BP; 3329 A; 3146 C; 3311 G; 3676 T; 0 U; 0 Other;
 SQ
 Query Match 62.0%; Score 73.8; DB 9; Length 13462;
 Best Local Similarity 87.1%; Pred. No. 9.4e-16;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

PT Novel nucleotide construct useful for disrupting function of gene in
PT embryonic stem cell, comprises sequence encoding positive selection
PT marker flanked by restriction enzyme sites, to create a single-stranded
PT region.

PS Claim 11; Fig 3B; 37pp; English.

XX The present invention relates to a new nucleotide construct comprising a
XX sequence encoding positive selection marker flanked by restriction enzyme
XX sites. The restriction enzyme sites are flanked by sequences which are
XX not complementary to each other and which do not include at least one
XX single-stranded regions are created at each sequence lacking at least one
XX nucleotide. The invention is useful for disrupting the function of a
XX target sequence or gene in a cell e.g. embryonic stem cell, by inserting
XX sequences homologous to the target gene flanked by the positive selection
XX marker, to produce a targeting construct, and introducing the targeting
XX construct into the cell to produce a homologous recombinant and thus the
XX function of the target gene or sequence is disrupted. The homologous
XX sequences are sequences flanking the site in the target gene that is to
XX be disrupted. The invention eliminates the need for the traditional
XX hybridisation isolation of a single genomic clone, restriction mapping of
XX the clone and multiple cloning steps. The method of the invention is fast
XX and efficiently generates nucleotide construct, and reduces the time
XX required for making a knock-out vector. Isolating an individual genomic
XX clone or mapping the restriction sites within the clone is not needed for
XX the method. The method is ligation independent cloning. The present
XX nucleic acid sequence represents the PDG4 plasmid vector sequence of the
XX invention

XX Sequence 6355 BP; 1531 A; 1670 C; 1655 G; 1499 T; 0 U; 0 Other;

Query Match 62.0%; Score 73.8; DB 6; Length 6355;
Best Local Similarity 87.1%; Pred. No. 7.8e-16;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA 86
DB 4625 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA 4684

QY 87 TTCGGGGCCCTTCGACCTCGAGCCCAATATGGGA 119

DB 4685 CTCGGGGCCCTTCGACCTCGAGCCCAATATGGGA 4717

RESULT 12

AAAS17144

ID AAAS17144 standard; DNA; 6355 BP.

XX AC AAAS17144;

XX 14-FEB-2002 (first entry)

XX Gene targeting vector pDG4.

XX pDG4; ds; retina-specific nuclear receptor; gene targeting;
XX lymphoid-specific GPCR; melanocyte stimulating hormone receptor;
XX magnesium-dependent protein phosphatase; transgenic animal;
XX chemokine receptor 1-like protein; cGMP phosphodiesterase;
XX sulfotransferase gene; tumour; cancer; retinal degeneration;
XX retinitis pigmentosa; Green fluorescent protein; GFP.

XX Escherichia coli.

XX Aequorea victoria.

XX Synthetic.

XX Chimeric.

XX WO200167855-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US008664.

XX

PR 16-MAR-2000; 2000US-0190348P.

PR 22-MAR-2000; 2000US-0191128P.

PR 22-MAR-2000; 2000US-0191129P.

PR 22-MAR-2000; 2000US-0191142P.

PR 22-MAR-2000; 2000US-0191235P.

PR 22-MAR-2000; 2000US-0191238P.

PR 22-MAR-2000; 2000US-0191240P.

PR 15-MAY-2000; 2000US-0204227P.

PR 29-JUN-2000; 2000US-0215214P.

PR 06-JUL-2000; 2000US-0216249P.

PR 06-JUL-2000; 2000US-0216264P.

PR 08-JUL-2000; 2000US-0216765P.

PR 12-JUL-2000; 2000US-0218075P.

PR 19-JUL-2000; 2000US-0219167P.

PR 19-JUL-2000; 2000US-0219182P.

PR 27-JUL-2000; 2000US-0221485P.

PR 07-AUG-2000; 2000US-0223173P.

XX (DELT-) DELTAGEN INC.

XX Allen KD, Guenther C, Phillips R;

XX WPI; 2002-041167/05.

XX New targeting construct comprising a first and a second polynucleotide

XX homologous to a target gene, and a selectable marker, useful for

XX introducing targeted mutations into embryonic cells.

XX Disclosure; Fig 3B; 105pp; English.

XX The invention relates to a targeting construct comprising two sequences

XX homologous to a target gene, and a selectable marker, is new. The target

XX gene is a retina-specific nuclear receptor gene, a lymphoid-specific GPCR

XX (G protein coupled receptor) gene, a melanocyte stimulating hormone

XX receptor gene, a magnesium-dependent protein phosphatase gene, chemokine

XX receptor 1-like protein gene, a cGMP phosphodiesterase gene, or a

XX sulfotransferase gene. Also included are transgenic mice comprising a

XX disruption in a target gene, where the mouse exhibits an eye abnormality,

XX cellular infiltration, hypoactive behaviour, lung abnormality, elevated

XX white blood cell count, abnormality in the aorta, kidney, liver, lymph

XX nodes, skin or salivary gland, increased body and organ weight, or

XX elevated levels of ALT (not defined), phosphorus, potassium, or

XX bilirubin, aggressive, hyperactive, increased activity or decreased

XX anxiety behaviour. The construct is used for introducing targeted

XX mutations into embryonic cells. The animal and cell-based systems may be

XX used as models for diseases or conditions associated with physiological,

XX histological or behavioural phenotypes relating to a disruption in a

XX target gene (e.g. tumours, cancer, retinal degeneration and retinitis

XX pigmentosa) and in screening or identifying compounds capable of

XX ameliorating or treating diseases. The present sequence is the vector

XX pDG4 used to generate the gene targeting construct of the invention.

XX pDG4 is similar to pDG2 (AAAS17143) except that it contains a gene for the

XX marker protein GFP (green fluorescent protein)

XX Sequence 6355 BP; 1535 A; 1668 C; 1657 G; 1495 T; 0 U; 0 Other;

XX Query Match 62.0%; Score 73.8; DB 6; Length 6355;

XX Best Local Similarity 87.1%; Pred. No. 7.8e-16;

XX Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA 86

DB 4625 CCGGCAATTCGCAAGCTTCAAAAGCGCAGCTCTGCGCGCTATTGTGAGCGCTCAAA 4684

QY 87 TTCGGGGCCCTTCGACCTCGAGCCCAATATGGGA 119

DB 4685 CTCGGGGCCCTTCGACCTCGAGCCCAATATGGGA 4717

RESULT 13

AAD28660

PT contains heterozygous disruption in a gene encoding TRP.
 XX
 PS Disclosure; Fig 3B; 106pp; English.
 XX
 CC The present sequence for plasmid vector pDG4 is used as a construct for
 CC genes encoding trinucleotide repeat proteins (TRP) such as gene T243 to
 CC produce disruption in the DNA. The invention describes methods of
 CC producing embryonic stem (ES) cells comprising a heterozygous disruption
 CC in a target DNA sequence (preferably T243) encoding a TRP and of
 CC producing a knockout mouse comprising a homozygous disruption in a gene
 CC encoding TRP, where the disruption inhibits the production of the wild
 CC type TRP. The invention also relates to identifying agents capable of
 CC affecting a phenotype of a knockout mouse. Also described are methods of
 CC determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and the
 CC cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse
 XX
 SQ Sequence 6355 BP; 1531 A; 1670 C; 1655 G; 1499 T; 0 U; 0 Other;
 Query Match 62.0%; Score 73.8; DB 4; Length 6355;
 Best Local Similarity 87.1%; Pred. No. 7.8e-16;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGACGCTCTCGCGGTATTGTGAGCGCTCACAA 86
 DB 4625 CCGGCAATTCGCAAGCTTCAAAAGCGACGCTCTCGCGGTATTGTCTCTCTCAT 4684
 QY 87 TTCGGGCGCTTTCGACCTGCGCAATATATGGGA 119
 DB 4685 CTCGGGCGCTTTCGACCTGCGCAATATATGGGA 4717
 RESULT 10
 ABL42020
 ID ABL42020 standard; DNA; 6355 BP.
 XX
 AC ABL42020;
 XX
 DT 11-JUN-2002 (first entry)
 XX
 DE Nucleotide sequence of vector pDG4.
 XX
 KW pDG4; transgenic animal; matrix metalloproteinase-23 gene; MMP-23 gene;
 KW ss.
 XX
 OS Synthetic.
 XX
 PN US2002023275-A1.
 XX
 PD 21-FEB-2002.
 XX
 PF 17-MAY-2001; 2001US-00861077.
 XX
 PR 17-MAY-2000; 2000US-0204972P.
 PR 29-JUN-2000; 2000US-0215394P.
 XX
 XX (LEVI/) LEVITEN M W.
 XX
 XX Leviten MW;
 XX
 XX WPI; 2002-255684/30.
 XX
 XX Non-human transgenic animal useful as a model for disease and for
 XX identifying agents that modulate gene expression and gene function.
 PT

PT comprises a disruption in the matrix metalloproteinase-23 gene.
 XX
 PS Disclosure; Fig 3B; 36pp; English.
 XX
 CC The present sequence represents vector pDG4. This vector contains an
 CC apicillin resistance gene, a neomycin gene and a green fluorescent
 CC protein (GFP) gene. The vector is used in the invention. The
 CC specification describes a non-human transgenic animal comprising a
 CC disruption in the matrix metalloproteinase (MMP)-23 gene. Transgenic
 CC animals of the invention comprising a homozygous or heterozygous
 CC disruption in MMP23 gene are useful for identifying agents which modulate
 CC MMP23 expression or function. They are also useful for identifying agents
 CC that are capable of ameliorating a phenotype of a transgenic animal
 CC comprising a disruption in an MMP-23 gene or ameliorating a disease
 CC associated with the phenotype of a transgenic animal comprising a
 CC disruption in the MMP-23 gene. The animals are useful as an animal model
 CC for diseases, disorders and conditions characterized by a disruption in a
 CC gene encoding a metalloproteinase, more particularly disease, disorders
 CC and conditions associated with the phenotypes demonstrated by the
 CC knockout mice. The transgenic animals are useful as test substrates for
 CC identification of drugs, pharmaceuticals and therapies effective in
 CC treating diseases, disorders and conditions associated with disruption in
 CC the target gene. The animal is useful for testing and developing new
 CC treatments relating to behavioural phenotypes demonstrated by the animal
 CC models
 XX
 SQ Sequence 6355 BP; 1535 A; 1668 C; 1657 G; 1495 T; 0 U; 0 Other;
 Query Match 62.0%; Score 73.8; DB 6; Length 6355;
 Best Local Similarity 87.1%; Pred. No. 7.8e-16;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 27 CCGGCAATTCGCAAGCTTCAAAAGCGACGCTCTCGCGGTATTGTGAGCGCTCACAA 86
 DB 4625 CCGGCAATTCGCAAGCTTCAAAAGCGACGCTCTCGCGGTATTGTCTCTCTCAT 4684
 QY 87 TTCGGGCGCTTTCGACCTGCGCAATATATGGGA 119
 DB 4685 CTCGGGCGCTTTCGACCTGCGCAATATATGGGA 4717
 RESULT 11
 ABS53352
 ID ABS53352 standard; DNA; 6355 BP.
 XX
 AC ABS53352;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Plasmid vector pDG4 DNA sequence.
 XX
 KW Nucleotide construct; positive selection marker; restriction enzyme site;
 KW embryonic stem cell; knock-out vector; genomic clone; mapping; PDG4;
 KW plasmid vector; ds.
 XX
 OS Synthetic.
 XX
 PN US2002086369-A1.
 XX
 PD 04-JUL-2002.
 XX
 PF 19-JUN-2001; 2001US-00885816.
 XX
 PR 17-NOV-1997; 97US-0084194P.
 PR 11-MAY-1998; 98US-0084949P.
 PR 17-NOV-1998; 98US-00193834.
 XX
 XX (DELT-) DELTAGEN INC.
 XX
 XX Klein RD, Brennan TJ;
 XX
 XX WPI; 2002-635678/68.
 XX
 XX

```
PA (DELT-) DELTAGEN INC.
XX
XX PI Allen KD, Guenther C, Phillips R;
XX
XX DR WPI; 2002-041167/05.
XX
XX PT New targeting construct comprising a first and a second polynucleotide
XX homologous to a target gene, and a selectable marker, useful for
XX introducing targeted mutations into embryonic cells.
XX
XX PS Example 3; Fig 2B; 105pp; English.
XX
XX CC The invention relates to a targeting construct comprising two sequences
XX homologous to a target gene, and a selectable marker, is new. The target
XX gene is a retina-specific nuclear receptor gene, a lymphoid-specific GPCR
XX (G protein coupled receptor) gene, a melanocyte stimulating hormone
XX receptor gene, a magnesium-dependent protein phosphatase gene, chemokine
XX receptor 1-like protein gene, a cAMP phosphodiesterase gene, or a
XX sulfoltransferase gene. Also included are transgenic mice comprising a
XX disruption in a target gene, where the mouse exhibits an eye abnormality,
XX cellular infiltration, hypoaffective behaviour, lung abnormality, elevated
XX white blood cell count, abnormality in the aorta, kidney, liver, lymph
XX nodes, skin or salivary gland, increased body and organ weight, or
XX elevated levels of ALT (not defined), phosphorus, potassium, or
XX bilirubin, aggressive, hyperactive, increased activity or decreased
XX anxiety behaviour. The construct is used for introducing targeted
XX mutations into embryonic cells. The animal and cell-based systems may be
XX used as models for diseases or conditions associated with physiological,
XX histological or behavioural phenotypes relating to a disruption in a
XX target gene (e.g. tumours, cancer, retinal degeneration and retinitis
XX pigmentosa) and in screening or identifying compounds capable of
XX ameliorating or treating diseases. The present sequence is the vector
XX pDG2 used to generate the gene targeting construct of the invention
XX
XX SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;
XX
XX Query Match 62.0%; Score 73.8; DB 6; Length 4768;
XX Best Local Similarity 87.1%; Pred. No. 7.2e-16;
XX Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
XX
XX QY 27 CCGCGCATTCGCGAGCTTCAAAAGCGCACGCTCTGCGCGCTATTGTGAGCGCTCACA 86
XX Db CCGCGCATTCGCGAGCTTCAAAAGCGCACGCTCTGCGCGCTATTGTCTCTCTCTCAT 3097
XX
XX QY 87 TTCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 119
XX Db CTCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 3130
XX
XX RESULT 8
XX AAD28659
XX ID AAD28659 standard; DNA; 4768 BP.
XX
XX AC AAD28659;
XX
XX DT 07-MAY-2002 (first entry)
XX
XX DE Plasmid pDG2 vector.
XX
XX KW Plasmid pDG2; DNA construct; embryonic stem cell; cell disruption; Neo';
XX KW neomycin; ampicillin resistance gene; ds.
XX
XX OS Unidentified.
XX
XX PN WO200204621-A2.
XX
XX XX 17-JAN-2002.
XX
XX XX 11-JUL-2000; 2000WO-US018812.
XX
XX PR 11-JUL-2000; 2000WO-US018812.
XX
XX PA (DELT-) DELTAGEN INC.
XX
XX PI Allen KD, Guenther C, Phillips R;
XX
XX DR WPI; 2002-041167/05.
XX
XX PT New targeting construct comprising a first and a second polynucleotide
XX homologous to a target gene, and a selectable marker, useful for
XX introducing targeted mutations into embryonic cells.
XX
XX PS Example 3; Fig 2B; 105pp; English.
XX
XX CC The invention relates to a targeting construct comprising two sequences
XX homologous to a target gene, and a selectable marker, is new. The target
XX gene is a retina-specific nuclear receptor gene, a lymphoid-specific GPCR
XX (G protein coupled receptor) gene, a melanocyte stimulating hormone
XX receptor gene, a magnesium-dependent protein phosphatase gene, chemokine
XX receptor 1-like protein gene, a cAMP phosphodiesterase gene, or a
XX sulfoltransferase gene. Also included are transgenic mice comprising a
XX disruption in a target gene, where the mouse exhibits an eye abnormality,
XX cellular infiltration, hypoaffective behaviour, lung abnormality, elevated
XX white blood cell count, abnormality in the aorta, kidney, liver, lymph
XX nodes, skin or salivary gland, increased body and organ weight, or
XX elevated levels of ALT (not defined), phosphorus, potassium, or
XX bilirubin, aggressive, hyperactive, increased activity or decreased
XX anxiety behaviour. The construct is used for introducing targeted
XX mutations into embryonic cells. The animal and cell-based systems may be
XX used as models for diseases or conditions associated with physiological,
XX histological or behavioural phenotypes relating to a disruption in a
XX target gene (e.g. tumours, cancer, retinal degeneration and retinitis
XX pigmentosa) and in screening or identifying compounds capable of
XX ameliorating or treating diseases. The present sequence is the vector
XX pDG2 used to generate the gene targeting construct of the invention
XX
XX SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;
XX
XX Query Match 62.0%; Score 73.8; DB 6; Length 4768;
XX Best Local Similarity 87.1%; Pred. No. 7.2e-16;
XX Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
XX
XX QY 27 CCGCGCATTCGCGAGCTTCAAAAGCGCACGCTCTGCGCGCTATTGTGAGCGCTCACA 86
XX Db CCGCGCATTCGCGAGCTTCAAAAGCGCACGCTCTGCGCGCTATTGTCTCTCTCTCAT 3097
XX
XX QY 87 TTCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 119
XX Db CTCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 3130
XX
XX RESULT 9
XX AAS05244
XX ID AAS05244 standard; DNA; 6355 BP.
XX
XX AC AAS05244;
XX
XX DT 07-SEP-2001 (first entry)
XX
XX DE Plasmid vector pDG4 used as a construct for TRP genes.
XX
XX KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES; pDG4;
XX KW transgenic animal; knockout mouse; triplet repeat expansion;
XX KW fragile X syndrome; Huntington's disease; cyclic; circular; ds.
XX
XX OS Synthetic.
XX
XX PN WO200130796-A1.
XX
XX PD 03-MAY-2001.
XX
XX PF 26-OCT-2000; 2000WO-US029382.
XX
XX PR 26-OCT-1999; 99US-0161488P.
XX
XX PA (DELT-) DELTAGEN INC.
XX
XX PI Klein R, Matthews W, Moore M, Allen KD;
XX
XX DR WPI; 2001-300473/31.
XX
XX PT Novel transgenic animals useful as animal model for characterization of
XX function of a gene encoding trinucleotide repeat proteins (TRPs),
```

CC disease, disorders and conditions associated with the phenotypes
 CC demonstrated by the knockout mice. The transgenic animals are useful as
 CC test substrates for identification of drugs, pharmaceuticals and
 CC therapies effective in treating diseases, disorders and conditions
 CC associated with disruption in the target gene. The animal is useful for
 CC testing and developing new treatments relating to behavioural phenotypes
 CC demonstrated by the animal models
 XX
 SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;

Query Match 62.0%; Score 73.8; DB 6; Length 4768;
 Best Local Similarity 87.1%; Pred. No. 7.2e-16;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGACGCTTCAAAAGCGACGCTCTCGCGGCTATTGTGAGCGCTCAAA 86
 DB 3038 CCGGCAATTCGACGCTTCAAAAGCGACGCTCTCGCGGCTATTGTGAGCGCTCAAT 3097
 QY 87 TTCGGGCTTTTCGACCTCGAGCAATATATGGA 119
 DB 3098 CTCGGGCTTTTCGACCTCGAGCAATATATGGA 3130

RESULT 6
 ABS53351
 ID ABS53351 standard; DNA; 4768 BP.

AC ABS53351;
 DT 15-NOV-2002 (first entry)
 DE Plasmid vector pDG2 DNA sequence.

Nucleotide construct; positive selection marker; restriction enzyme site;
 embryonic stem cell; knock-out vector; genomic clone; mapping; pDG2;
 plasmid vector; ds.
 Synthetic.

US2002086369-A1.

04-JUL-2002.

19-JUN-2001; 2001US-00885816.

17-NOV-1997; 97US-0084194P.

11-MAY-1998; 98US-0084949P.

17-NOV-1998; 98US-00193834.

(DELT-) DELTAGEN INC.

Klein RD, Brennan TJ;

WPI; 2002-635678/68.

Novel nucleotide construct useful for disrupting function of gene in
 embryonic stem cell, comprises sequence encoding positive selection
 marker flanked by restriction enzyme sites, to create a single-stranded
 region.

Claim 10; Fig 2A; 37pp; English.

The present invention relates to a new nucleotide construct comprising a
 sequence encoding positive selection marker flanked by restriction enzyme
 sites. The restriction enzyme sites are flanked by sequences which are
 not complementary to each other and which do not include at least one
 type of base at any position, where the construct is treated so that
 single-stranded regions are created at each sequence lacking at least one
 nucleotide. The invention is useful for disrupting the function of a
 target sequence or gene in a cell e.g. embryonic stem cell, by inserting
 sequences homologous to the target gene into the construct such that the
 sequences homologous to the target gene flank the positive selection
 marker, to produce a targeting construct, and introducing the targeting

CC construct into the cell to produce a homologous recombinant and thus the
 CC function of the target gene or sequence is disrupted. The homologous
 CC sequences are sequences flanking the site in the target gene that is to
 CC be disrupted. The invention eliminates the need for the traditional
 CC hybridisation isolation of a single genomic clone, restriction mapping of
 CC the clone and multiple cloning steps. The method of the invention is fast
 CC and efficiently generates nucleotide construct, and reduces the time
 CC required for making a knock-out vector. Isolating an individual genomic
 CC clone or mapping the restriction sites within the clone is not needed for
 CC the method. The method is ligation independent cloning. The present
 CC nucleic acid sequence represents the pDG2 plasmid vector sequence of the
 CC invention

SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;

Query Match 62.0%; Score 73.8; DB 6; Length 4768;

Best Local Similarity 87.1%; Pred. No. 7.2e-16;

Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGCAATTCGACGCTTCAAAAGCGACGCTCTCGCGGCTATTGTGAGCGCTCAAA 86

DB 3038 CCGGCAATTCGACGCTTCAAAAGCGACGCTCTCGCGGCTATTGTGAGCGCTCAAT 3097

QY 87 TTCGGGCTTTTCGACCTCGAGCAATATATGGA 119

DB 3098 CTCGGGCTTTTCGACCTCGAGCAATATATGGA 3130

RESULT 7

AAS17143

ID AAS17143 standard; DNA; 4768 BP.

AC AAS17143;

DT 14-FEB-2002 (first entry)

DE Gene targeting vector pDG2.

KW pDG2; ds; retina-specific nuclear receptor; gene targeting;

lymphoid-specific GPCR; melanocyte stimulating hormone receptor;

magnesium-dependent protein phosphatase; transgenic animal;

chemokine receptor 1-like protein; cGMP phosphodiesterase;

sulfotransferase gene; tumour; cancer; retinal degeneration;

retinitis pigmentosa.

OS Escherichia coli.

OS Synthetic.

XX WO200167855-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US008664.

XX 16-MAR-2000; 2000US-0190348P.

XX 22-MAR-2000; 2000US-0191128P.

XX 22-MAR-2000; 2000US-0191128P.

XX 22-MAR-2000; 2000US-0191142P.

XX 22-MAR-2000; 2000US-0191235P.

XX 22-MAR-2000; 2000US-0191235P.

XX 22-MAR-2000; 2000US-0191240P.

XX 15-MAY-2000; 2000US-0204227P.

XX 15-MAY-2000; 2000US-0204230P.

XX 29-JUN-2000; 2000US-0215214P.

XX 06-JUL-2000; 2000US-0216249P.

XX 06-JUL-2000; 2000US-0216264P.

XX 12-JUL-2000; 2000US-0218075P.

XX 19-JUL-2000; 2000US-0219167P.

XX 19-JUL-2000; 2000US-0219182P.

XX 27-JUL-2000; 2000US-0221485P.

XX 07-AUG-2000; 2000US-0223173P.

XX

CC a cell with high efficiency and specificity. Use of (I) provides a faster
CC and more efficient means for isolating and selecting cells comprising
CC target gene modifications. Also use of (I) provides an increase over
CC previous technologies in both the speed and frequency at which homologous
CC recombination events can be recovered. (I) is also useful for creation of
CC transgenic animals containing targeted gene modifications. This sequence
CC represents a phosphoglycerate kinase (PGK)-neomycin (neo) gene fragment
CC found in the positive selection vector construct c3406
XX
SQ Sequence 108 BP; 17 A; 41 C; 25 G; 25 T; 0 U; 0 Other;

Query Match 62.0%; Score 73.8; DB 6; Length 108;
Best Local Similarity 87.1%; Pred. No. 2.7e-16;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCGCAAGCTTCAAAAGCGCACGCTCGCGGCTATTGTGAGGCGCTCACAA 86

Db 16 CCGGGCATTCGCAAGCTTCAAAAGCGCACGCTCGCGGCTATTGTGAGGCGCTCACAT 75

QY 87 TTCGGGCGCTTCGACCTCGAGCAATATGGGA 119

Db 76 CTCGGGCGCTTCGACCTCGAGCAATATGGGA 108

RESULT 4
AAS05243
ID AAS05243 standard; DNA; 4768 BP.

XX AAS05243;

XX 07-SEP-2001 (first entry)

XX Plasmid vector pDG2 used as a construct for TRP genes.

XX Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES; pDG2;
XX transgenic animal; knockout mouse; triplet repeat expansion;
XX fragile X syndrome; Huntington's disease; cyclic; circular; ds.

XX Synthetic.

XX WO200130798-A1.

XX 03-MAY-2001.

XX 26-OCT-2000; 2000WO-US029382.

XX 26-OCT-1999; 99US-0161488P.

XX (DELT-) DELTAGEN INC.

XX Klein R, Matthews W, Moore M, Allen KD;

XX WPI; 2001-300473/31.

XX Novel transgenic animals useful as animal model for characterization of
PT function of a gene encoding trinucleotide repeat proteins (TRPs),
PT contains heterozygous disruption in a gene encoding TRP.

XX Disclosure; Fig 2B; 106pp; English.

XX The present sequence for plasmid vector pDG2 is used as a construct for
CC genes encoding trinucleotide repeat proteins (TRP) such as gene T243 to
CC produce disruption in the DNA. The invention describes methods of
CC producing embryonic stem (ES) cells comprising a heterozygous disruption
CC in a target DNA sequence (preferably T243) encoding a TRP and of
CC producing a knockout mouse comprising a homozygous disruption in a gene
CC encoding TRP, where the disruption inhibits the production of the wild
CC type TRP. The invention also relates to identifying agents capable of
CC affecting a phenotype of a knockout mouse. Also described are methods of
CC determining whether expansion of the trinucleotide repeat in a gene
CC encoding TRP produces a phenotypic change. The transgenic animals and the
CC cells are useful for identifying compounds capable of ameliorating
CC disease symptoms, and as test substrates for the identification of drugs,

CC pharmaceuticals, therapies and interventions which may be effective in
CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
CC Huntington's disease. The animal models for trinucleotide repeat
CC disorders are ideal model systems to study the progression of disease in
CC vivo, the molecular basis of these diseases and show the features
CC observed in human disease. Using the mice, it is possible to model both
CC the pathogenic mechanism and the trinucleotide repeat instability in the
CC mouse
XX
SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 U; 0 Other;

Query Match 62.0%; Score 73.8; DB 4; Length 4768;
Best Local Similarity 87.1%; Pred. No. 7.2e-16;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 27 CCGGGCATTCGCAAGCTTCAAAAGCGCACGCTCGCGGCTATTGTGAGGCGCTCACAA 86

Db 3038 CCGGGCATTCGCAAGCTTCAAAAGCGCACGCTCGCGGCTATTGTGAGGCGCTCACAT 3097

QY 87 TTCGGGCGCTTCGACCTCGAGCAATATGGGA 119

Db 3098 CTCGGGCGCTTCGACCTCGAGCAATATGGGA 3130

RESULT 5
ABL42019
ID ABL42019 standard; DNA; 4768 BP.

XX ABL42019;

XX 11-JUN-2002 (first entry)

XX Nucleotide sequence of vector pDG2.

XX pDG2; transgenic animal; matrix metalloproteinase-23 gene; MMP-23 gene;
XX ss.

XX Synthetic.

XX US2002023275-A1.

XX 21-FEB-2002.

XX 17-MAY-2001; 2001US-00861077.

XX 17-MAY-2000; 2000US-0204972P.

XX 29-JUN-2000; 2000US-0215394P.

XX (LEVI/) LEVITEN M W.

XX Leviten MW;

XX WPI; 2002-255684/30.

XX Non-human transgenic animal useful as a model for disease and for
PT identifying agents that modulate gene expression and gene function,
PT comprises a disruption in the matrix metalloproteinase-23 gene.

XX Example 1; Fig 2B; 38pp; English.

XX The present sequence represents vector pDG2. This vector contains an
CC ampicillin resistance gene and a neomycin gene. The vector is used in the
CC invention. The specification describes a non-human transgenic animal
CC comprising a disruption in the matrix metalloproteinase (MMP)-23 gene.
CC Transgenic animals of the invention comprising a homozygous or
CC heterozygous disruption in MMP23 gene are useful for identifying agents
CC which modulate MMP23 expression or function. They are also useful for
CC identifying agents that are capable of ameliorating a phenotype of a
CC transgenic animal comprising a disruption in an MMP-23 gene or
CC ameliorating a disease associated with the phenotype of a transgenic
CC animal comprising a disruption in the MMP-23 gene. The animals are useful
CC as an animal model for diseases, disorders and conditions characterized
CC by a disruption in a gene encoding a metalloproteinase, more particularly

CC cells having targeting vector integrated via homologous recombination
 CC into the genomes of the cells. (I) is capable of modifying target gene in
 CC a cell with high efficiency and specificity. Use of (I) provides a faster
 CC and more efficient means for isolating and selecting cells comprising
 CC target gene modification. Also use of (I) provides an increase over
 CC previous technologies in both the speed and frequency at which homologous
 CC recombination events can be recovered. (I) is also useful for creation of
 CC transgenic animals containing targeted gene modifications. This sequence
 CC represents a partial phosphoglycerate kinase (PGK) promoter incorporated
 CC into the positive selection vector constructs described in the invention
 XX
 XX Sequence 119 BP; 26 A; 37 C; 29 G; 27 T; 0 U; 0 Other;

Query Match 100.0%; Score 119; DB 6; Length 119;
 Best Local Similarity 100.0%; Pred. No. 2.8e-32;
 Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGTCTCTATTGTGAGCGCTCACAAATCCGGCATCTCTCGAAGCTTCAAAGCGCACGTC 60
 DB 1 AAGTCTCTATTGTGAGCGCTCACAAATCCGGCATCTCTCGAAGCTTCAAAGCGCACGTC 60
 QY 61 TGGCGCGCTATTGTGAGCGCTCACAAATTCGGGCGCTTTCGACCTGCGAGCCAAATATGGGA 119
 DB 61 TGGCGCGCTATTGTGAGCGCTCACAAATTCGGGCGCTTTCGACCTGCGAGCCAAATATGGGA 119

RESULT 2
 ABK49513
 ID ABK49513 standard; DNA; 75 BP.

XX AC ABK49513;

XX DT 15-JUL-2002 (first entry)

XX DE Positive selection vector associated oligonucleotide 10218.

XX KW Transgenic animal; targeting vector; positive selection vector;

XX KW homologous recombination; target gene modification; transgenic animal;

XX KW ss.

XX OS Synthetic.

XX PN WO200222834-A2.

XX PD 21-MAR-2002.

XX PF 17-SEP-2001; 2001WO-US028892.

XX PR 15-SEP-2000; 2000US-0232957P.

XX PA (DELT-) DELTAGEN INC.

XX PI Siebel C, Brennan TJ;

XX DR WPI; 2002-383132/41.

XX Novel targeting vector modifying target gene, has first and second
 PT sequences homologous to target gene portions, a selectable marker
 PT cassette and regulator, useful for producing animals with targeted gene
 PT modifications.

XX Example 1; Fig 8; 43pp; English.

XX The invention describes a targeting vector (positive selection vector)
 CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
 CC and S2) homologous to a portion or region of a target gene, a selectable
 CC marker cassette and a regulator. (I) is useful for producing cells
 CC comprising a modification of the target gene which involves introducing
 CC (I) into cells capable of homologous recombination, selecting for cells
 CC expressing the selectable marker and identifying cells containing the
 CC modification of the target gene. Use of (I) for enriching cells
 CC comprising disruption or modification of target gene enhances recovery of
 CC cells having targeting vector integrated via homologous recombination

CC into the genomes of the cells. (I) is capable of modifying target gene in
 CC a cell with high efficiency and specificity. Use of (I) provides a faster
 CC and more efficient means for isolating and selecting cells comprising
 CC target gene modification. Also use of (I) provides an increase over
 CC previous technologies in both the speed and frequency at which homologous
 CC recombination events can be recovered. (I) is also useful for creation of
 CC transgenic animals containing targeted gene modifications. This sequence
 CC represents an oligonucleotide used in the creation of the positive
 CC selection vectors for targeted gene modification
 XX
 XX Sequence 75 BP; 14 A; 25 C; 18 G; 18 T; 0 U; 0 Other;

Query Match 63.0%; Score 75; DB 6; Length 75;
 Best Local Similarity 100.0%; Pred. No. 9.4e-17;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 GCATTCTCGAAGCTTCAAAGCGCACGTCGCGCGCTATTGTGAGCGCTCACAAATCC 90
 DB 1 GCATTCTCGAAGCTTCAAAGCGCACGTCGCGCGCTATTGTGAGCGCTCACAAATCC 60
 QY 91 GGGCCTTTGACCTG 105
 DB 61 GGGCCTTTGACCTG 75

RESULT 3

ABK49508

ID ABK49508 standard; DNA; 108 BP.

XX AC ABK49508;

XX DT 15-JUL-2002 (first entry)

XX DE Phosphoglycerate kinase (PGK)-neomycin (neo) gene fragment.

XX KW Transgenic animal; targeting vector; positive selection vector;

XX KW homologous recombination; target gene modification; transgenic animal;

XX KW phosphoglycerate kinase; PGK; neomycin; neo; ds.

XX OS Synthetic.

XX PN WO200222834-A2.

XX PD 21-MAR-2002.

XX PF 17-SEP-2001; 2001WO-US028892.

XX PR 15-SEP-2000; 2000US-0232957P.

XX PA (DELT-) DELTAGEN INC.

XX PI Siebel C, Brennan TJ;

XX DR WPI; 2002-383132/41.

XX Novel targeting vector modifying target gene, has first and second
 PT sequences homologous to target gene portions, a selectable marker
 PT cassette and regulator, useful for producing animals with targeted gene
 PT modifications.

XX Example 1; Fig 6C; 43pp; English.

XX The invention describes a targeting vector (positive selection vector)
 CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
 CC and S2) homologous to a portion or region of a target gene, a selectable
 CC marker cassette and a regulator. (I) is useful for producing cells
 CC comprising a modification of the target gene which involves introducing
 CC (I) into cells capable of homologous recombination, selecting for cells
 CC expressing the selectable marker and identifying cells containing the
 CC modification of the target gene. Use of (I) for enriching cells
 CC comprising disruption or modification of target gene enhances recovery of
 CC cells having targeting vector integrated via homologous recombination
 CC into the genomes of the cells. (I) is capable of modifying target gene in

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:24:05 ; Search time 103.493 Seconds
(without alignments)
4884.713 Million cell updates/sec

Title: US-09-954-483B-2
Perfect score: 119
Sequence: 1 aaggtctattgtgagcgt.....gacctgagccaataatggga 119

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	119	100.0	119	6	ABK49509
2	75	63.0	75	6	ABK49513
3	73.8	62.0	108	6	ABK49508
4	73.8	62.0	4768	4	AAS05243
5	73.8	62.0	4768	6	ABL42019
6	73.8	62.0	4768	6	ABSS3351
7	73.8	62.0	4768	6	AAS17143
8	73.8	62.0	4768	6	AAD28659
9	73.8	62.0	6355	4	AAS05244
10	73.8	62.0	6355	6	ABL42020
11	73.8	62.0	6355	6	ABSS3352
12	73.8	62.0	6355	6	AAS17144
13	73.8	62.0	6355	6	AAD28660
14	73.8	62.0	13462	9	ADB68453
15	64.2	53.9	6505	8	ADA12886
16	63.4	53.3	8934	7	AAL41385
17	62.6	52.6	10491	9	ADB81348
18	62.6	52.6	12538	9	ADB81342
19	62.6	52.6	12645	9	ADB81343
20	59.2	49.7	7090	9	AAD20746
21	57.8	48.6	513	5	AAL20729
22	56.8	47.7	11784	6	ABT08201
23	56.8	47.7	11784	9	ADB81350

C	24	56	47.1	66	6	ABK49515
	25	54.6	45.9	5365	5	AAD04928
	26	54.2	45.5	4847	4	AAD09280
	27	53.8	45.2	15692	2	AAX24731
	28	53.8	45.2	15692	2	AAX24732
C	29	53.8	45.2	15701	2	AAX24733
	30	53.6	45.0	3997	6	ABQ78682
	31	53.6	45.0	4430	9	ABQ78682
	32	53.6	45.0	7523	6	ADD13783
	33	53.6	45.0	7608	6	ABT08174
	34	53.6	45.0	7803	6	ABT08173
	35	53.6	45.0	8167	6	ABT08177
	36	53	44.5	5377	3	AA53872
	37	53	44.5	9725	3	AA53873
	38	53	44.5	9732	3	AA53879
	39	53	44.5	9873	3	AA53875
	40	53	44.5	10054	3	AA53876
C	41	52.8	44.4	7617	2	AAV14354
	42	52.2	43.9	5581	5	AAL41035
	43	52	43.7	4083	9	ADD13934
C	44	52	43.7	4953	8	ADA94772
	45	50.6	42.5	17135	7	AAL41384

ALIGNMENTS

RESULT 1
ABK49509
ID ABK49509 standard; DNA; 119 BP.
XX
AC ABK49509;
XX
DT 15-JUL-2002 (first entry)
XX
DE Partial Phosphoglycerate kinase (PGK) promoter.
XX
KW Transgenic animal; targeting vector; positive selection vector;
KW homologous recombination; target gene modification; transgenic animal;
KW phosphoglycerate kinase; PGK; promoter; ds.
XX
OS Unidentified.
XX
PN WO200222834-A2.
XX
PD 21-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-US028892.
XX
PR 15-SEP-2000; 2000US-0232957P.
XX
PI (DELT-) DELTAGEN INC.
XX
PI Siebel C, Brennan TJ;
XX
WI MPI; 2002-383132/41.
XX
PT Novel targeting vector modifying target gene, has first and second
PT sequences homologous to target gene portions, a selectable marker
PT cassette and regulator, useful for producing animals with targeted gene
PT modifications.
XX
PS Claim 9; Fig 6D; 43pp; English.
XX
CC The invention describes a targeting vector (positive selection vector)
CC (I) capable of modifying a target gene. (I) comprises two sequences (S1
CC and S2) homologous to a portion or region of a target gene, a selectable
CC marker cassette and a regulator. (I) is useful for producing cells
CC comprising a modification of the target gene which involves introducing
CC (I) into cells capable of homologous recombination, selecting for cells
CC expressing the selectable marker and identifying cells containing the
CC modification of the target gene. Use of (I) for enriching cells
CC comprising disruption or modification of target gene enhances recovery of

```

source      1..5608
            /organism="Cloning vector pNTKV1903"
            /mol_type="genomic DNA"
            /db_xref="taxon:82684"
            /lab_host="Escherichia coli K12"

ORIGIN
Query Match      53.1%; Score 63.2; DB 12; Length 5608;
Best Local Similarity 85.4%; Pred. No. 2.1e-10;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY      24 AATCCCGGCATTCTCGCAAGCTTCAAAGCGCACGTCTGCCGCGCTATTGTGAGCGCTCA 83
      |||||
Db      3285 AAGCCCGGCATTCT-SCACGCTTCAAAGCGCACGTCTGCCGCGCTATTGTCTCTCTCT 3227

QY      84 CAATTCCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 119
      |||||
Db      3226 CATCTCCGGGCGCTTTCGACCTGCAGCCCAATATGGGA 3191

Search completed: April 22, 2004, 11:46:17
Job time : 669.215 secs

```

```

/organism="Cloning vector pXONEO"
/mol_type="genomic DNA"
/db_xref="taxon:82121"
/lab_host="Escherichia coli K12"
/plasmid="pXONEO"
/notes="part of the knockout mouse system produced by
Lexicon"

ORIGIN
Query Match      53.1%; Score 63.2; DB 12; Length 3426;
Best Local Similarity 85.4%; Pred. No. 2.1e-10;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 24 AATCCGGCATTCTCGAAGCTTCAAAGCGACGCTCTGCCGCGCTATTGTGAGCGCTCA 83
Db 2237 AAGCCGGCATTCT-GCAGCTTCAAAGCGACGCTCTGCCGCGCTATTCTCTCTCT 2295

QY 84 CAATTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 119
Db 2296 CATCTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 2331

RESULT 12
AF092169/c
LOCUS AF092169 5608 bp DNA circular SYN 04-OCT-1998
DEFINITION Cloning vector pNTKV1907, complete sequence.
ACCESSION AF092169
VERSION AF092169.1 GI:3694963
KEYWORDS
SOURCE Cloning vector pNTKV1907
ORGANISM Cloning vector pNTKV1907
REFERENCE 1 (bases 1 to 5608)
AUTHORS Skinner,H. and Pearce,M.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1998) Technical Services, Stratagene Cloning
Systems, 11011 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
source
1..5608
/organism="Cloning vector pNTKV1907"
/mol_type="genomic DNA"
/db_xref="taxon:82682"
/lab_host="Escherichia coli K12"

ORIGIN
Query Match      53.1%; Score 63.2; DB 12; Length 5608;
Best Local Similarity 85.4%; Pred. No. 2.1e-10;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 24 AATCCGGCATTCTCGAAGCTTCAAAGCGACGCTCTGCCGCGCTATTGTGAGCGCTCA 83
Db 3287 AAGCCGGCATTCT-GCAGCTTCAAAGCGACGCTCTGCCGCGCTATTCTCTCTCT 3229

QY 84 CAATTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 119
Db 3228 CATCTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 3193

RESULT 13
AF092172/c
LOCUS AF092172 5608 bp DNA circular SYN 04-OCT-1998
DEFINITION Cloning vector pNTKV1901, complete sequence.
ACCESSION AF092172
VERSION AF092172.1 GI:3694964
KEYWORDS
SOURCE Cloning vector pNTKV1901
ORGANISM Cloning vector pNTKV1901
REFERENCE 1 (bases 1 to 5608)
AUTHORS Skinner,H. and Grafsky,A.J.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Technical Services, Stratagene, 11011 N.
Torrey Pines Rd., La Jolla, CA 92037, USA

```

```

FEATURES
source
1..5608
/organism="Cloning vector pNTKV1901"
/mol_type="genomic DNA"
/db_xref="taxon:82679"
/lab_host="Escherichia coli K12"

ORIGIN
Query Match      53.1%; Score 63.2; DB 12; Length 5608;
Best Local Similarity 85.4%; Pred. No. 2.1e-10;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 24 AATCCGGCATTCTCGAAGCTTCAAAGCGACGCTCTGCCGCGCTATTGTGAGCGCTCA 83
Db 3286 AAGCCGGCATTCT-GCAGCTTCAAAGCGACGCTCTGCCGCGCTATTCTCTCTCT 3228

QY 84 CAATTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 119
Db 3227 CATCTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 3192

RESULT 14
AF092173/c
LOCUS AF092173 5608 bp DNA circular SYN 04-OCT-1998
DEFINITION Cloning vector pNTKV1902, complete sequence.
ACCESSION AF092173
VERSION AF092173.1 GI:3694965
KEYWORDS
SOURCE Cloning vector pNTKV1902
ORGANISM Cloning vector pNTKV1902
REFERENCE 1 (bases 1 to 5608)
AUTHORS Skinner,H. and Grafsky,A.J.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Technical Services, Stratagene, 11011 N.
Torrey Pines Rd., La Jolla, CA 92037, USA

FEATURES
source
1..5608
/organism="Cloning vector pNTKV1902"
/mol_type="genomic DNA"
/db_xref="taxon:82683"
/lab_host="Escherichia coli K12"

ORIGIN
Query Match      53.1%; Score 63.2; DB 12; Length 5608;
Best Local Similarity 85.4%; Pred. No. 2.1e-10;
Matches 82; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 24 AATCCGGCATTCTCGAAGCTTCAAAGCGACGCTCTGCCGCGCTATTGTGAGCGCTCA 83
Db 3287 AAGCCGGCATTCT-GCAGCTTCAAAGCGACGCTCTGCCGCGCTATTCTCTCTCT 3229

QY 84 CAATTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 119
Db 3228 CATCTCCGGGCTTTGACCTCGACCTGAGCAATATGGGA 3193

RESULT 15
AF092174/c
LOCUS AF092174 5608 bp DNA circular SYN 04-OCT-1998
DEFINITION Cloning vector pNTKV1903, complete sequence.
ACCESSION AF092174
VERSION AF092174.1 GI:3694966
KEYWORDS
SOURCE Cloning vector pNTKV1903
ORGANISM Cloning vector pNTKV1903
REFERENCE 1 (bases 1 to 5608)
AUTHORS Skinner,H. and Grafsky,A.J.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Technical Services, Stratagene, 11011 N.
Torrey Pines Rd., La Jolla, CA 92037, USA

FEATURES
Location/Qualifiers

```

```

QY      87  TTCGGGCGCTTTCGACCTCGACCGCAATATGGGA 119
Db      4685  CTCGGGCGCTTTCGACCTCGACCGCAATATGGGA 4717

RESULT 8
AX352705
LOCUS      AX352705                6355 bp      DNA      linear      PAT 06-FEB-2002
DEFINITION Sequence 2 from Patent WO0204621.
ACCESSION  AX352705
VERSION     AX352705.1  GI:18617826
KEYWORDS   .
SOURCE      synthetic construct
            synthetic construct
            artificial sequences.
ORGANISM    1
REFERENCE   1
AUTHORS     Klein,R.D. and Brennan,T.J.
TITLE       Methods of creating constructs useful for introducing sequences in
            to embryonic stem cells
JOURNAL     Patent: WO 0204621-A 2 17-JAN-2002;
            Deltagen, Inc. (US)
FEATURES    Location/Qualifiers
             source
               1..6355
               /organism="synthetic construct"
               /mol_type="unassigned DNA"
               /db_xref="taxon:32630"
               /note="Plasmid vector"

ORIGIN
Query Match      62.0%; Score 73.8; DB 6; Length 6355;
Best Local Similarity 87.1%; Pred. No. 5.6e-14;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      27  CCCGGCATTTCGCAAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTGAGCGCTCACAA 86
Db      4625  CCCGGCATTTCGCAAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTCTCTCTCAT 4684

QY      87  TTCGGGCGCTTTCGACCTCGACCGCAATATGGGA 119
Db      4685  CTCGGGCGCTTTCGACCTCGACCGCAATATGGGA 4717

RESULT 9
AX823827
LOCUS      AX823827                6505 bp      DNA      linear      PAT 12-DEC-2003
DEFINITION Sequence 5 from Patent WO03070958.
ACCESSION  AX823827
VERSION     AX823827.1  GI:39750146
KEYWORDS   .
SOURCE      murine stem cell retroviral vector (MSCV)
            murine stem cell retroviral vector
            artificial sequences; vectors.
ORGANISM    1
REFERENCE   1
AUTHORS     Yao,S. and Ellis,J.
TITLE       Retroviral gene therapy vectors including insulator elements to
            provide high levels of gene expression
JOURNAL     Patent: WO 03070958-A 5 28-AUG-2003;
            THE HOSPITAL FOR SICK CHILDREN (CA)
FEATURES    Location/Qualifiers
             source
               1..6505
               /organism="murine stem cell retroviral vector"
               /mol_type="unassigned DNA"
               /db_xref="taxon:258023"
               1..6505
               /note="n = g, a, t, or c."

LTR

ORIGIN
Query Match      53.9%; Score 64.2; DB 6; Length 6505;
Best Local Similarity 77.2%; Pred. No. 9.8e-11;
Matches 78; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY      19  CTCACAATCCCGGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTGAGC 78
Db      4685  CTCACAATCCCGGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTGAGC 4717

RESULT 10
AX593541
LOCUS      AX593541                8934 bp      DNA      linear      PAT 13-FEB-2003
DEFINITION Sequence 56 from Patent WO02088353.
ACCESSION  AX593541
VERSION     AX593541.1  GI:28374904
KEYWORDS   .
SOURCE      synthetic construct
            synthetic construct
            artificial sequences.
ORGANISM    1
REFERENCE   1
AUTHORS     Chambon,P., Ghyselinck,N.B. and Schnuettgen,F.
TITLE       Method for the stable inversion of dna sequence by site-specific
            recombination and dna vectors and transgenic cells thereof
JOURNAL     Patent: WO 02088353-A 56 07-NOV-2002;
            ADEREGEM (FR)
FEATURES    Location/Qualifiers
             source
               1..8934
               /organism="synthetic construct"
               /mol_type="unassigned DNA"
               /db_xref="taxon:32630"
               /note="DNA sequence of plasmid pJM8-Position 1 to 2240
               vector sequence-Position 2245 Frt site, sense-Position
               2285 loxP1 site, sense-Position 2355 lox511 site,
               sense-Position 2400 to 5952 NLS-Lac2 polyA gene,
               antisense-Position 5960 to 6549 IRES, antisense-Position
               6550 to 7050 rabbit beta globin intron, antisense-Position
               7060 loxP1 site, antisense-Position 7115 to 7630 FGK
               promoter, sense-Position 7638 to 8840 Neomycine resistance
               gene, sense-Position 8441 to 8480 synthetic splice donor
               site, sense-Position 8505 lox511 site, antisense-Position
               8540 Frtm site, antisense-Position 8600 to 8934 vector
               sequence"

ORIGIN
Query Match      53.3%; Score 63.4; DB 6; Length 8934;
Best Local Similarity 83.2%; Pred. No. 1.8e-10;
Matches 84; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY      19  CTCACAATCCCGGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTGAGC 78
Db      7544  CTCGGAGCGCGCGCATTTCT-GCAGCTTCAAAGCGCAGCTCTGCCGCGTATTGTCTCTC 7602

QY      79  GCTCACAATTCGGGGCGCTTTTCGACCTCGACCGCAATATGGGA 119
Db      7603  TTCCTCATCTCCGGGCGCTTTTCGACCTCGACCGCAATATGGGA 7643

RESULT 11
AF090454
LOCUS      AF090454                3426 bp      DNA      circular SYN 28-AUG-2002
DEFINITION Cloning vector pKONEO complete sequence.
ACCESSION  AF090454
VERSION     AF090454.1  GI:3608512
KEYWORDS   .
SOURCE      Cloning vector pKONEO
            Cloning vector pKONEO
            artificial sequences; vectors.
ORGANISM    1 (bases 1 to 3426)
REFERENCE   1
AUTHORS     Skinner,H.
TITLE       Direct Submission
JOURNAL     Submitted (08-SEP-1998) Lexicon Genetics Inc., 400 Research Forrest
            Dr., The Woodlands, TX 77381, USA
FEATURES    Location/Qualifiers
             source
               1..3426

```

```
JOURNAL Patent: WO 0204621-A 1 17-JAN-2002;
Deltagen, Inc. (US)
FEATURES
  source
    1..4768
      /location/Qualifiers
        /organism="synthetic construct"
        /mol_type="unassigned DNA"
        /db_xref="taxon:32630"
        /note="Plasmid vector"
ORIGIN
Query Match 62.0%; Score 73.8; DB 6; Length 4768;
Best Local Similarity 87.1%; Pred. No. 5.6e-14;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCAAA 86
Db 3038 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCA 3097
QY 87 TTCGCGGCTTTCGACCTGCGACGCAATATGGGA 119
Db 3098 CTCGCGGCTTTCGACCTGCGACGCAATATGGGA 3130
RESULT 6
AF397196/c 6293 bp DNA circular SYN 21-AUG-2001
LOCUS Retrofitting vector pRetroS, complete sequence.
DEFINITION Retrofitting vector pRetroS, complete sequence.
ACCESSION AF397196
VERSION AF397196.1 GI:15216973
KEYWORDS Retrofitting vector pRetroS
SOURCE Retrofitting vector pRetroS
ORGANISM Retrofitting vector pRetroS
REFERENCE Wang, Z., Engler, P., Longacre, A. and Storb, U.
TITLE An efficient method for high-fidelity BAC/PAC retrofitting with a
selectable marker for mammalian cell transfection
Genome Res. 11 (1), 137-142 (2001)
JOURNAL MEDLINE
PUBMED 21086695
REFERENCE Wang, Z., Engler, P., Longacre, A. and Storb, U.
AUTHORS Direct Submission
TITLE Submitted (05-JUL-2001) Mol. Genet. Cell Biol., University of
Chicago, 920 E. 58th St., Chicago, IL 60637, USA
JOURNAL Location/Qualifiers
FEATURES
  source
    1..6293
      /organism="Retrofitting vector pRetroS"
      /mol_type="genomic DNA"
      /db_xref="taxon:168860"
      /note="for in vivo retrofitting of BAC/PAC clones"
      /gene="cre"
      561..2363
      /gene="cre"
      561..2363
      /gene="cre"
      /codon_start=1
      /transl_table=11
      /product="GST-loxp-cre recombinase fusion protein"
      /protein_id="AAK92454.1"
      /db_xref="GI:15216974"
      /translation="MGYWKIGKLVQPTLLILEEYKEEHLIERDSDGKWRKKFE
      LGLEFNPYYIDGVDYKQSMALIRYIADKNMILGGCPKRAEISMLEGAVLDIRY
      VRIAYSKDFETLKVDFLSKPELMKMFEDRLCHTKYTLNGDHVTHPDFMLYDALDVL
      YMDMCLDAPFKLVEFKTEIAPQIDKYLKSKYIAPWLOGQATFGGQHPKPSDL
      VPRGSRAVGHMPMAITSYHYTKLSGIPRAEHMANALLTVHONLEALPVDAI
      SDVRKNLMDFRDRAQSEHTWMLLSVCRSWAAWCKLNKKWFPAPEDVRDYLLY
      LGAGLVKTIQOHLGOLANKLHRSGLPEPSNAVSLVMBIRIKENVDAGERAKQAL
      AFERTDFQVRSIMENSDQDIRNLAFIAGIAYNTLLRIAEIARIRVKDISRTDGRM
      LIHGKTKIVSTAGVKAUSLVGKVERWISVSGVADDPNNILFCRVKRGVAAPS
      MQAGGWTNIVNMYIRNLDSETGMVRLLEDGD"
      1332..2360
      /gene="cre"
misc_feature
```

```

  CDS
    complement(3064..3867)
      /note="Region: cre recombinase"
      /note="for mammalian cell selection"
      /codon_start=1
      /transl_table=11
      /product="PGK-neo"
      /protein_id="AAK92455.1"
      /db_xref="GI:15216975"
      /translation="MGSATIEDQDLHAGSPAAWVERLFGYDWAQQTIGSDAAVRLSA
      QGRPVLVFKTDLSGALNEQDEARLSWLTATGVCFAVLDVVTEAGRMDWLLGEVPG
      QQLSSHLAPAEKSVIMADAMERLHTLPATCFPDHQAHRTERATMEAGLVQDD
      LBERGCLPAEFAELKAKMPDGDLLVTHSDACLPMIWMVNGFSFGFIDCGRLGVA
      DRYQDIALARDIAELGGEWADRFVLVINGIAAPDSQRIAFRLLEDF"
      4486..4905
      /note="RK conditional replication origin"
      /complement(5032..5892)
      /note="ampicillin-resistance protein"
      /codon_start=1
      /transl_table=11
      /product="beta-lactamase"
      /protein_id="AAK92456.1"
      /db_xref="GI:15216976"
      /translation="MSIQHFRVALIPFFAFCIPVFAHPETLVKVKDAEDQLGARVGY
      IELDLSNGKILSPRPEPSPMWTFKVLGGVLSRVDAQGEQLGRIRHYSQNDLVE
      YSPVTEKHLTDGMTVRELCSAAITMSDNTAALLTTIGPKELTAFLHMGDHTVRL
      DRWPELNAIPNDERDTTPVAMATLKLTGELLTLASRQQLIDWMEADKVAQPL
      LRSALPAGWFIADKSGAGERGRIALGPDGKFSRIIVITIGSQATMDENRQIA
      EIGASLIKHW"
ORIGIN
Query Match 62.0%; Score 73.8; DB 12; Length 6293;
Best Local Similarity 87.1%; Pred. No. 5.6e-14;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCAAA 86
Db 3954 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCAT 3895
QY 87 TTCGCGGCTTTCGACCTGCGACGCAATATGGGA 119
Db 3994 CTCGCGGCTTTCGACCTGCGACGCAATATGGGA 3862
RESULT 7
LOCUS AX299822
DEFINITION Sequence 2 from Patent WO0167855.
ACCESSION AX299822
VERSION AX299822.1 GI:17129313
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Allen, K.D., Guenther, C. and Phillips, R.
TITLE Transgenic mice containing targeted gene disruptions
JOURNAL Patent: WO 0167855-A 2 20-SEP-2001;
Deltagen, Inc. (US)
FEATURES
  source
    1..6355
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Phage Vector"
ORIGIN
Query Match 62.0%; Score 73.8; DB 6; Length 6355;
Best Local Similarity 87.1%; Pred. No. 5.6e-14;
Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCAAA 86
Db 4625 CCCGCAATTCGCAAGCTTCAAAAGCGACGCTGCGCGCTATTGAGCGCTCAT 4684
```

```
JOURNAL Patent: WO 0222834-A 1 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..119
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Targeting Vector"
ORIGIN
  Query Match 100.0%; Score 119; DB 6; Length 119;
  Best Local Similarity 100.0%; Pred. No. 2.9e-29;
  Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTCTATTGTGAGCGCTCACAAATCCGGCAATTCGCAAGCTTCAAAGCGCAGTC 60
DB 1 AAGGTCTATTGTGAGCGCTCACAAATCCGGCAATTCGCAAGCTTCAAAGCGCAGTC 60
QY 61 TGCCGCGCTATTGTGAGCGCTCACAAATTCGCGGCTTTCGACCTGCAGCCAAATATGGGA 119
DB 61 TGCCGCGCTATTGTGAGCGCTCACAAATTCGCGGCTTTCGACCTGCAGCCAAATATGGGA 119
RESULT 2
AX468463
LOCUS 75 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 6 from Patent WO0222834.
ACCESSION AX468463
VERSION AX468463.1 GI:21901299
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
        artificial sequences.
REFERENCE
  1 Siebel, C. and Brennan, T.J.
  Methods of producing cells and animals comprising targeted gene
  modifications
  JOURNAL Patent: WO 0222834-A 6 21-MAR-2002;
  Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..75
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Primer"
ORIGIN
  Query Match 63.0%; Score 75; DB 6; Length 75;
  Best Local Similarity 100.0%; Pred. No. 2.1e-14;
  Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 GCATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAAATCC 90
DB 1 GCATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAAATCC 60
QY 91 GGGCCTTTCGACCTG 105
DB 61 GGGCCTTTCGACCTG 75
RESULT 3
AX468458
LOCUS 108 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0222834.
ACCESSION AX468458
VERSION AX468458.1 GI:21901294
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
        artificial sequences.
REFERENCE
  1 Siebel, C. and Brennan, T.J.
  Methods of producing cells and animals comprising targeted gene
  modifications
JOURNAL Patent: WO 0222834-A 1 21-MAR-2002;
Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..108
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Targeting Vector"
ORIGIN
  Query Match 62.0%; Score 73.8; DB 6; Length 108;
  Best Local Similarity 87.1%; Pred. No. 5.4e-14;
  Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 CCCGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAA 86
DB 16 CCCGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAA 75
QY 87 TTCCGGGCGCTTTCGACCTGCAGCCAAATATGGGA 119
DB 76 TTCCGGGCGCTTTCGACCTGCAGCCAAATATGGGA 108
RESULT 4
AX299821
LOCUS 4768 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 1 from Patent WO0167855.
ACCESSION AX299821
VERSION AX299821.1 GI:17129312
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
        artificial sequences.
REFERENCE
  1 Allen, K.D., Guenther, C. and Phillips, R.
  Transgenic mice containing targeted gene disruptions
  JOURNAL Patent: WO 0167855-A 1 20-SEP-2001;
  Deltagen, Inc. (US)
FEATURES
  source
    Location/Qualifiers
      1..4768
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Phage Vector"
ORIGIN
  Query Match 62.0%; Score 73.8; DB 6; Length 4768;
  Best Local Similarity 87.1%; Pred. No. 5.4e-14;
  Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 CCCGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAA 86
DB 3038 CCCGCGATTCTCGCAAGCTTCAAAGCGCAGCTCTCGCGGCTATTGTGAGCGCTCACAA 3097
QY 87 TTCCGGGCGCTTTCGACCTGCAGCCAAATATGGGA 119
DB 3098 TTCCGGGCGCTTTCGACCTGCAGCCAAATATGGGA 3130
RESULT 5
AX352704
LOCUS 4768 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0204621.
ACCESSION AX352704
VERSION AX352704.1 GI:18617825
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
        artificial sequences.
REFERENCE
  1 Klein, R.D. and Brennan, T.J.
  Methods of creating constructs useful for introducing sequences in
  to embryonic stem cells
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 10:25:20 ; Search time 665.215 Seconds
(without alignments)
7753.612 Million cell updates/sec

Title: US-09-954-483B-2
Perfect score: 119
Sequence: 1 aaggtccatctgtgagcgt.....gacotgagccaatatggga 119

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.ste.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	100.0	119	6	AX468459	Sequence
2	75	63.0	75	6	AX468463	Sequence
3	73.8	62.0	108	6	AX468458	Sequence
4	73.8	62.0	4768	6	AX299821	Sequence
5	73.8	62.0	4768	6	AX352704	Sequence
6	73.8	62.0	6293	12	AF397196	Sequence
7	73.8	62.0	6355	6	AX299822	Sequence
8	73.8	62.0	6355	6	AX352705	Sequence
9	64.2	53.9	6505	6	AX23827	Sequence
10	63.4	53.3	8934	6	AX593541	Sequence
11	63.2	53.1	3426	12	AF090454	Cloning v
12	63.2	53.1	5608	12	AF092169	Cloning v
13	63.2	53.1	5608	12	AF092172	Cloning v
14	63.2	53.1	5608	12	AF092173	Cloning v
15	63.2	53.1	5608	12	AF092174	Cloning v
16	63.2	53.1	5608	12	AF092541	Cloning v
17	63.2	53.1	5608	12	AF092542	Cloning v
18	63.2	53.1	5608	12	AF092543	Cloning v
19	63.2	53.1	5608	12	AF092567	Cloning v
20	62.8	52.8	14863	12	AF339820	Cloning v
21	62.6	52.6	5588	12	AF335419	Cloning v
22	62.6	52.6	5588	12	AF335420	Cloning v
23	62.6	52.6	10491	6	AX816389	Sequence
24	62.6	52.6	12538	6	AX816383	Sequence
25	62.6	52.6	12645	6	AX816384	Sequence
26	59.2	49.7	7090	6	AR310582	Sequence
27	59.2	49.7	7090	6	AX150263	Sequence
28	58.8	49.4	1110	10	MUSPGK1	Mouse phosph
29	57.8	48.6	513	6	AR310565	Sequence
30	57.8	48.6	513	6	AX150246	Sequence
31	57.2	48.1	5830	12	AY028413	YTT vecto
32	56.8	47.7	11784	6	AX816391	Sequence
33	56.8	47.7	22960	12	AY192024	BAC cloni
34	56	47.1	66	6	AX468465	Sequence
35	55.8	46.9	5842	12	AY028415	YTT vecto
36	55.4	46.6	8451	12	PVFM3BL	Plasmid vec
37	54.6	45.9	5365	6	AX114854	Sequence
38	54.2	45.5	4847	6	AX191674	Sequence
39	53.6	45.0	3997	6	AX589681	Sequence
40	53.6	45.0	4106	6	AX777461	Sequence
41	53.6	45.0	4430	6	AX777463	Sequence
42	53	44.5	3297	12	AF090453	Cloning v
43	53	44.5	4699	12	AF346823	RAGE vect
44	53	44.5	5382	6	AR215117	Sequence
45	53	44.5	5382	6	AR302359	Sequence

ALIGNMENTS

RESULT 1
AX468459
LOCUS AX468459
DEFINITION Sequence 2 from Patent WO222834.
ACCESSION AX468459
VERSION AX468459.1 GI:21901295
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Siebel, C. and Brennan, T.J.
TITLE Methods of producing cells and animals comprising targeted gene
modifications
JOURNAL Patent: WO 0222834-A 2 21-MAR-2002;

DNA linear PAT 16-JUL-2002

09/954,483
3 MAY 2004
Examiner's Search Notes